C9orf72 frontotemporal dementia and amyotrophic lateral sclerosis: investigating repeat pathology in cell culture models and human post-mortem brain.

Charlotte Elizabeth Ridler

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy from University College London

Department of Neurodegenerative Disease

Institute of Neurology

University College London

2016

Declaration

I, Charlotte Ridler, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

My deepest thanks go to everyone who has helped and supported me during this PhD project, without whom it would not have been possible. Thank you firstly to my supervisor Dr. Adrian Isaacs for all of his guidance, encouragement and enthusiastic discussions during this exciting and unpredictable project. Thank you also to Dr. Sarah Mizielinska, who has been an incredible mentor and role-model during this PhD; I have learnt a huge amount from her endless dedication and generosity. I would also like to thank my secondary supervisor Prof. Elizabeth Fisher for all her wise words of advice.

Thank you to all members of the Isaacs and Fisher labs past and present. Special thanks to Frances Norona, Dr. Karen Cleverly, Julian Pietrzyk, Dr. Ione Woollacott and Dr. Anny Devoy who were all great help during the cloning phase of this project; to Dr. Rubika Balendra for all her help and discussions in the nucleoli work; and to Dr. Emma Clayton and Dr. Roberto Simone for their constant support.

A number of other people have lent me their valuable expertise for this project, for which I am very grateful. Thank you to Dr. Tammaryn Lashley for provision of histological samples and for her valuable discussions, to Dr. Teresa Nicolli for all her work and discussions regarding the fly project, and to Dr. Vincent Plagnol for his vital statistics advice.

I am also extremely grateful to my funders Alzheimer's Research UK for their support for this project.

Thank you to fellow PhDs Dr. Xun Yu Choong and Angelos Armen with whom I have shared the entire rollercoaster ride of the last three years. Thank you also to everyone in the PhD office for all your support and shenanigans, James Miller, Laura Pulford, Justin Tosh, Ludmila Sheytanova, Dr. Billy West, Lucianne Dobson, Dr. Rachelle Saccon, Amy Nick, Julian Jaeger, Ruchi Kumari, and Alexandra Philiastides. Thank you also to all of the support staff at the Institute of Neurology.

Finally, I would like to thank all of my wonderful family and friends outside of the lab who supported me, encouraged me and made me cups of tea: my partner Thomas Jennings, my parents Kay and Tony Ridler, my brother Samuel Ridler, all of my extended family, my supportive dance friends, and my housemates Thomas Beales Ferguson, Craig Buchanan, Jacob Spencer, Abigail Attrell, Adam Ladds, Oliver Mattos, and Otso Liikanen.

Abstract

A GGGGCC repeat expansion in the first intron of chromosome 9 open reading fame 72 (*C9orf72*) is the most common known genetic cause of both frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS). The repeats are transcribed into RNA in both sense and antisense orientations, which aggregates to form RNA foci in cells. Additionally, repeat RNA undergoes repeat-associated non-ATG initiated (RAN) translation producing dipeptide repeat (DPR) proteins in all six sense and antisense reading frames. This thesis aims to dissect RNA and DPR protein gain-of-function mechanisms operating in *C9orf72* FTD/ALS to study their effects in isolation. To achieve this, three classes of DNA constructs were generated: 1) pure GGGGCC repeats that produce RNA foci and DPR proteins, 2) 'RNA-only' constructs designed to preclude RAN translation, and 3) 'protein-only' constructs coding for all five DPR proteins that do not produce GGGGCC repeat RNA. Cultured cells transfected with pure and RNA-only repeat constructs exhibited a positive correlation between repeat length and both sense and antisense RNA foci formation. Additionally, DPR protein subcellular localisation was examined in protein-only repeat construct transfected cells and compared to the pathology found in patient brain. A collaborative project expressing these constructs in *Drosophila* showed that arginine-containing DPR proteins were responsible for *C9orf72* repeat toxicity in flies. To determine the relevance of the different DPR proteins to human disease, the frequency and distribution of DPR protein inclusions was studied in *C9orf72* FTD patient brain, using a novel automated image analysis protocol. Poly(glycine-arginine) DPR protein inclusion-bearing neurons also exhibited an increased nucleolar volume compared to inclusion-negative neurons, confirming recent speculation implicating nucleolar stress in disease pathogenesis. These studies in cell culture, Drosophila and human post-mortem brain implicate poly(glycine-arginine) as a toxic species within *C9orf72* FTD, which may aid in the targeting of future treatments for this condition.

Table of Contents

List of figures

List of Tables

Abbreviations

AAV	Adeno-associated virus
AD	Alzheimer's Disease
ADARB2	Adenosine deaminase RNA specific B2
AGD	Argyrophilic grain disease
ALS	Amyotrophic lateral sclerosis
ALS-FUS	Amyotrophic lateral sclerosis with FUS inclusions
ALS-SOD1	Amyotrophic lateral sclerosis with SOD1 inclusions
ALS-TDP	Amyotrophic lateral sclerosis with TDP-43 inclusions
ANOVA	Analysis of variance
ASO	Antisense Oligonucleotide
Aβ	Amyloid-beta
BAC	Bacterial artificial chromosome
BMAA	beta-Methylamino-L-alanine
bp	Base pair
BSA	Bovine serum albumen
bvFTD	Behavioural variant frontotemporal dementia
C9orf72	Chromosome 9 open reading frame 72
CBS	Corticobasal syndrome
CD	Circular dichroism
CIP	Calf intestinal phosphatase
CpG	Cytosine-phosphate-guanine
CSF	Cerebrospinal fluid
Ctrl	Control
DAB	3,3-Diaminobenzidine
DAPI	4',6-diamidino-2-phenylindole
ddH_2O	Double-distilled water
DENN	Differentially expressed in normal and neoplasia
DM1	Myotonic dystrophy type 1
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPR	Dipeptide repeat
EDTA	Ethylenediaminetetraacetic acid
FCS	Fetal calf serum
FISH	Fluorescence *in situ* hybridisation
FMRP	Fragile X mental retardation protein
FTD	Frontotemporal dementia
FTD/ALS	Frontotemporal dementia and/or amyotrophic lateral sclerosis
FTD-ALS	Combined frontotemporal dementia and amyotrophic lateral sclerosis
FTLD	Frontotemporal lobar degeneration
FTLD-FUS	Frontotemporal lobar degeneration with FUS inclusions

FTLD-tau	Frontotemporal lobar degeneration with tau pathology
FTLD-TDP	Frontotemporal lobar degeneration with TDP-43 inclusions
FTLD-UPS	Frontotemporal lobar degeneration with ubiquitinated inclusions
FUS	Fused in sarcoma
FXTAS	Fragile X-associated tremor/ataxia syndrome
GA	Glycine-alanine
GCL	Granule cell layer
GDP	Guanosine diphosphate
GEF	Guanine exchange factor
GFP	Green fluorescent protein
GP	Glycine-proline
GR	Glycine-arginine
GRN	Progranulin
GTP	Guanosine triphosphate
GWAS	Genome wide association study
hnRNP	Heterogeneous ribonucleoprotein particle
IBMPFD	Inclusion body myopathy with Paget's disease of bone and frontotemporal dementia
IF	Immunofluorescence
IHC	Immunohistochemistry
iPSC	Induced pluripotent stem cell
kb	Kilobase
LB	Luria-broth
LPH	Limulus polyphemus hemocyanin
MAPT	Microtubule-associated protein tau
MBNL1	Muscleblind-like protein 1
MCS	Multiple cloning site
ML	Molecular layer
MND	Motor neuron disease
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleicacid
NOR	Nucleolar organising region
NPC	Nuclear pore complex
PA	Proline-alanine
PAGE	Polyacrylamide gel electrophoresis
PBP	Progressive bulbar palsy
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PD	Parkinson's disease
PET	Positron emission tomography
PFA	Paraformaldehyde
PLS	Primary lateral sclerosis
PMA	Progressive muscular atrophy
PNFA	Progressive non-fluent aphasia
PPA	Primary progressive aphasia

PR	Proline-arginine
PSP	Progressive supranuclear palsy
RAN translation	Repeat-associated non-ATG initiated translation
RanGAP	Ran GTPase-activating protein
RDL	Recursive directional ligation
RNA	Ribonucleic acid
ROI	Region of interest
RRM	RNA recognition motif
SCC	Saline sodium citrate
SD	Semantic dementia
SEM	Standard error of mean
siRNA	Short interfacing RNA
snoRNP	Small nucleolar ribonuclear protein
SOC	Super optimal broth with catabolite repression
SOD1	Superoxide dismutase 1
TDP-43	Transactive response DNA binding protein 43
UTR	Untranslated region
VCP	Valosin containing protein

Chapter 1: Introduction

1.1 Overview

Neurodegeneration has a major impact on quality of life for both patients and those around them, and additionally presents an immense economic cost to society in terms of health and social care. Advancing age is the biggest risk factor for neurodegenerative disease, therefore as the global population grows and the proportion of older individuals increases, the number of patients suffering from neurodegenerative diseases is predicted to dramatically rise in the coming decades, with dementia cases in the UK predicted to double by 2050 (Lewis et al., 2014).

Our understanding of the genetics and molecular mechanisms behind neurodegeneration has greatly increased over the last 25 years. However, available treatment options remain limited: current therapies may offer some alleviation of symptoms, but do very little to delay disease progression. A more detailed knowledge of pathological mechanisms involved in neurodegeneration is therefore necessary to develop more effective therapies.

In 2011, two international collaborations resulted in the identification of a mutation in the gene *C9orf72* that causes two neurodegenerative disorders: frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS) (DeJesus-Hernandez et al., 2011; Renton et al., 2011b). This genetic defect is now considered to be the most common known cause of both FTD and ALS in most European and North American populations (Majounie et al., 2012b), however, how it gives rise to these diseases is not yet fully understood. The work in this thesis therefore explores the molecular mechanisms underlying *C9orf72* FTD/ALS to better understand the disease processes operating in these patients, which may also aid in our understanding of neurodegeneration as a whole.

1.2 Clinical features of frontotemporal lobar degeneration and amyotrophic lateral sclerosis

1.2.1 Frontotemporal dementia

Frontotemporal dementia is the most common clinical presentation of a neurodegenerative pathology termed frontotemporal lobar degeneration (FTLD), which is characterised by the progressive atrophy of the frontal and temporal cortices. The prevalence of FTD is estimated to be 15 cases per 100,000 in patients between 45 and 64 years of age, accounting for between 5 and 15 % of all dementia onset in patients younger than 65, rendering it the second most common form of pre-senile dementia after Alzheimer's disease (Rademakers et al., 2012; Ratnavalli et al., 2002). Incidence is equally distributed between men and women (Neary et al., 2005). Disease onset most often occurs between 45 and 65 years of age, with a median duration of 6-8 years, but can range from 2 to 20 years (Neary et al., 2005). Genetic factors can play a strong role in the disease, with a familial history of FTD or a related disorder present in 25 to 50 % of cases, while the remainder of cases are classed as sporadic (Rohrer et al., 2009; Seelaar et al., 2008). FTD can be subdivided into several clinical subgroups (Figure 1.1) including behavioural variant frontotemporal dementia (bvFTD) and two forms of language variants collectively known as primary progressive aphasia (PPA): semantic dementia (SD) and progressive non-fluent aphasia (PNFA) (Neary et al., 1998; Sieben et al., 2012). bvFTD is characterised by pronounced degeneration in the orbital and mesial frontal lobes, leading to changes in personality and social behaviour, loss of empathy, stereotyped behaviour and a lack of insight. Patients can present as socially disinhibited and inappropriate, but also as apathetic and unemotional. SD results from degeneration of the anterior temporal lobes, and results in a loss of conceptual knowledge and impaired comprehension and recollection of words, names, objects and faces. Patients increasingly substitute more general terms such as

"thing" in place of specific concepts. PNFA also results in a loss of language abilities, arising from degeneration of the inferior and insular frontal cortex, but while conceptual understanding of words and objects is retained, grammar and speech articulation is disrupted, leading to mispronunciation of words and disruption of sentence structure. Patients can also present with a mixed clinical syndrome of these subtypes, and aspects of both behavioural and language deficits often occur in the advanced stages of disease (Neary et al., 1998; Sieben et al., 2012). In addition, some overlap with corticobasal syndrome (CBS) and progressive supranuclear palsy (PSP) has been noted in FTD, resulting in a presentation of parkinsonism in some patients (Kertesz and Munoz, 2004). However, unlike the dementia exhibited in Alzheimer's disease, memory functions in FTD cases are usually left intact.

1.2.2 Amyotrophic lateral sclerosis

Motor neuron disease (MND) is a group of conditions involving progressive motor neuron loss, the most prevalent of which is amyotrophic lateral sclerosis (ALS), which accounts for more than 75 % of all MND cases (Figure 1.1). Prevalence is approximately 5.2 cases per 100,000, with men 1.2 to 1.5 times more likely than women to be diagnosed (Logroscino et al., 2010). ALS is characterised by the degeneration of both the upper motor neurons within layer V of the motor cortex, and lower motor neurons in the brainstem and anterior horn of the spinal cord. Clinically, this manifests as fasciculations, spasticity, and progressive wasting and weakness within voluntary muscles, most frequently beginning with symptoms in the limb muscles (65 %), but can also present with bulbar dysfunction (30 %), while a minority of cases experience respiratory onset (5%) (Logroscino et al., 2010). Other forms of MND include progressive muscular atrophy (PMA), in which degeneration occurs solely in the lower motor neurons, primary lateral schlerosis (PLS) in which only the upper motor neurons are affected, and progressive bulbar palsy (PBP) which principally affects the bulbar nuclei.

ALS disease course is frequently very rapid, but can vary considerably. Disease duration is less than 3 years in over 60 % of patients, while 10 % of patients survive more than 8 years, and rare cases have survived 40 years (Kiernan et al., 2011). 5 to 10 % of patients have a familial history of the disease, with the remaining cases arising sporadically. The average age of disease onset is between 47 to 52 years for familial ALS and 58 to 63 years for sporadic ALS (Kiernan et al., 2011). A number of environmental factors have been implicated in a higher risk of ALS onset, including smoking, high levels of physical activity, and exposure to pesticides and electromagnetic radiation (Alonso et al., 2010; Sutedja et al., 2009b), Additionally, a neurotoxic amino acid beta-methylamino-L-alanine (BMAA) produced by cyanobacteria is hypothesised to cause ALS when ingested in high quantities, leading to a high incidence of ALS in parts of the world where it enters the food chain (Bradley and Mash, 2009). A higher incidence of ALS has also been noted in certain professions including military workers, soccer players and health workers (Sutedja et al., 2009a), however larger population based studies are required to confirm these associations.

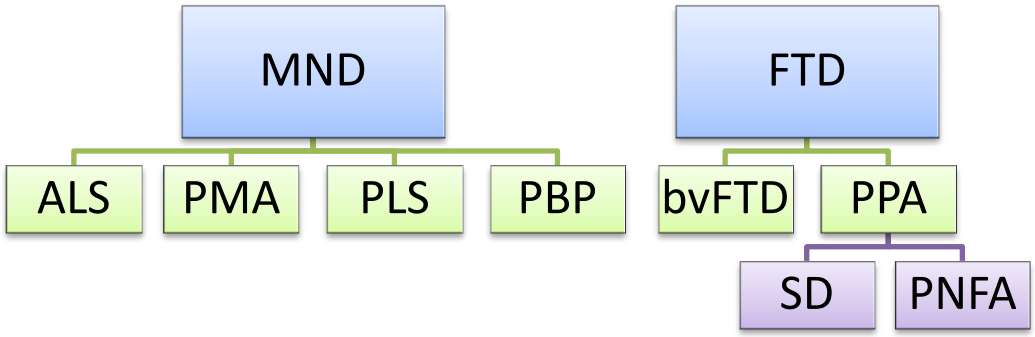


Figure 1.1 Clinical subtypes of frontotemporal lobar degeneration and motor neuron disease

FTD: frontotemporal dementia, MND: motor neuron disease, ALS: amyotrophic lateral sclerosis, PMA: progressive muscular atrophy, PLS: primary lateral sclerosis, PBP: progressive bulbar atrophy, bvFTD: behavioural variant frontotemporal dementia, PPA: primary progressive aphasia, SD: semantic dementia, PNFA: progressive non-fluent aphasia

1.2.3 Clinical overlap

While FTD and ALS were historically considered as separate entities, it has been increasingly recognised that these two neurodegenerative disorders are related, and in fact form part of a single pathological continuum (Lillo and Hodges, 2009). As many as half of all ALS patients exhibit some degree of behavioural or cognitive impairment, while the severity of these symptoms is severe enough to warrant a classification of FTD in 15 to 18 % of patients (Lomen-Hoerth et al., 2003; Ringholz et al., 2005). Additionally, around 40 % of FTD patients exhibit mild motor system dysfunction such as muscle wasting, muscle weakness and infrequent fasciculations, with 12 to 15 % of FTD patients exhibiting motor dysfunction that meets the criteria for a diagnosis of ALS (Burrell et al., 2011; Lomen-Hoerth et al., 2002). Motor neuron deficits observed in combined FTD-ALS most often resemble classical ALS, while cognitive changes observed in FTD-ALS are typically akin to bvFTD, with rare cases of SD or PNFA with ALS also reported (Van Langenhove et al., 2012). FTD-ALS patients often have a poor prognosis, with an average survival 2-3 years following symptom onset (Sieben et al., 2012). Heritability of FTD-ALS is also high, with 50 % considered familial, suggesting a strong genetic component (Sieben et al., 2012). In accordance with this clinical overlap, recent studies have increasingly uncovered common neuropathology (Section 1.3) and genetics (Section 1.4) underlying the coincidence of these disorders.

1.2.4 Treatment

Currently, the only treatment for ALS is riluzole, an inhibitor of presynaptic glutamate release, which prolongs median survival in patients by a modest two to three months and yields a 9 % increase in patient survival for one year (Miller et al., 2007). Regrettably, there are presently no disease modifying therapies available for FTD. Disease management usually consists of symptom treatment only, with a focus on palliative care, building

support networks and psychiatric services. Some Alzheimer's disease treatments for improving cognitive impairment have been tested in FTD patients, but little benefit has been observed so far, with some interventions even observed to worsen behavioural symptoms (Vossel and Miller, 2008).

1.3 Pathological features of FTLD and ALS

FTLD refers to the pathological processes that underlie the clinical presentation of FTD. Both FTLD and ALS pathologies are characterised by the formation of aberrant protein aggregates within neurons and glia, and pathological subtypes can be classified according to the primary protein components present within these aggregates. FTLD cases can be classified under four main categories: FTLD-tau, FTLD-TDP, FTLD-FUS and FTLD-UPS, while ALS cases can be classified under three main categories: ALS-TDP, ALS-SOD1 and ALS-FUS.

1.3.1 FTLD-TDP and ALS-TDP

In 2006, transactive response (TAR) DNA-binding protein 43 (TDP-43) was discovered to be the major pathological protein in the majority of tau-negative, SOD1-negative, ubiquitin-positive inclusions in both FTLD and ALS, accounting for approximately 50 % of all FTLD cases and over 95 % of all ALS cases (Mackenzie et al., 2007; Neumann et al., 2006). In addition, mutations in the *TARDBP* gene cause approximately 3 % of familial ALS and rare cases of FTD (Lattante et al., 2013). TDP-43 is a highly evolutionarily conserved member of the heterogeneous ribonucleoprotein (hnRNP) family, with roles in splicing regulation and transcription, which it mediates via its two RNA recognition motifs (RRMs) and its glycine-rich low-complexity c-terminal region (Mackenzie et al., 2010b). TDP-43 has been reported to bind over 6000 RNA targets in the brain, often targeting transcripts with very long introns and GU enriched sequences (Polymenidou et al., 2011). Under normal conditions, TDP-43 is primarily localised to the nucleus, but shuttles between the nucleus and the cytoplasm. However, in some forms of ALS and FTLD, TDP-43 is depleted from the nucleus

and forms abnormally phosphorylated and ubiquitylated aggregates in the cytoplasm, with truncated C-terminal fragments of the protein also observed (Mackenzie et al., 2010b). It is not yet clear how TDP-43 mislocalisation might lead to neurodegeneration, but it is thought to be due to a combination of loss of protein function in the nucleus, and gain of toxic function of the cytoplasmic aggregates. A loss of function mechanism is supported by studies in human cell lines in which a loss of viability was observed following knock-down of TDP-43 (Ayala et al., 2008; Iguchi et al., 2009), as well as in *Drosophila* and mouse models, in which motor defects were reported in response to reduced TDP-43 expression (Feiguin et al., 2009; Kraemer et al., 2010). Over expression of human TDP-43 was also shown to result in TDP-43 cytoplasmic inclusions and neurodegeneration in *Drosophila*, mouse and rat models (Li et al., 2010; Tatom et al., 2009; Wils et al., 2010).

In ALS cases, neuronal cytoplasmic inclusions of TDP-43 are observed primarily within neurons and glia within the motor cortex, spinal cord, and brainstem motor nuclei, as well as the white matter tracts associated with these areas (Mackenzie et al., 2007). Within FTLD, the distribution of TDP-43 pathology differs between patients and between different clinical or genetic subtypes. FTLD-TDP pathology can be categorised into four broad profiles: A, B, C and D (Mackenzie et al., 2011). Type A is characterised by frequent neuronal cytoplasmic inclusions and short dystrophic neurites positive for TDP-43, primarily within layer 2 of the neocortex. Type B TDP-43 pathology is exemplified by moderate neuronal cytoplasmic inclusions across all cortical layers, with rare dystrophic neurites. TDP-43 type C pathology is characterised by frequent long dystrophic neurites with infrequent neuronal cytoplasmic inclusions, predominantly within layer 2 of the neocortex. Finally, type D pathology is exhibited by patients with inclusion body myopathy with Paget's disease of bone and frontotemporal dementia (IBMPFD), presenting as frequent short dystrophic neurites, frequent neuronal intranuclear inclusions, and rare neuronal cytoplasmic inclusions, throughout all cortical layers.

1.3.2 FTLD-FUS and ALS-FUS

Fused in sarcoma (FUS) was identified as the primary component of the majority of non-TDP-43, ubiquitinated inclusions in FTLD and ALS, accounting for approximately 9 % of all cases of FTLD, and rare cases (less than 1 %) of ALS (Kwiatkowski et al., 2009; Vance et al., 2009). Conversely, mutations in FUS are observed to cause around 4 % of familial ALS cases, and rare cases of FTLD (Van Langenhove et al., 2010). Previously identified as a fusion oncogene, FUS is a highly conserved RNA and DNA binding protein with multiple binding domains including an RNA recognition motif, a transcriptional activation domain, a zinc finger motif and several glycine-arginine-rich "RGG" nucleic acid-binding domains (Deng et al., 2014). While the precise function of FUS has not been elucidated, it is known to be involved in transcriptional regulation, and has also been implicated in DNA damage repair (Deng et al., 2014). In a similar manner to TDP-43, FUS predominantly localises to the nucleus under normal conditions and shuttles between the nucleus and the cytoplasm, but forms frequent cytoplasmic and infrequent intranuclear inclusions under pathological conditions (Deng et al., 2014). Unlike TDP-43, aberrant phosphorylation or cleavage of FUS is not observed. FUS pathology has additionally been reported in other neurodegenerative disorders including many forms of polyglutamine disease (Doi et al., 2008).

1.3.3 ALS-SOD1

Superoxide dismutase 1 (SOD1) is an anti-oxidant enzyme whose principal function is the conversion of reactive oxygen species to hydrogen peroxide and oxygen, protecting the cell against oxidative stress. Mutations in SOD1 were first reported to cause ALS in 1993 (Rosen et al., 1993), and it was later found that misfolded SOD1 forms cytoplasmic aggregates in these patients (Kato et al., 2000). Patients with SOD1 mutations are now known to account for around 20 % of familial ALS cases around 1 % of sporadic ALS cases (Pasinelli and Brown, 2006). For over a decade, mutations in SOD1 remained the only known cause of

ALS, and were therefore studied intensely, with numerous transgenic mouse models generated, however the mechanism by which SOD1 causes ALS still remains elusive (McGoldrick et al., 2013). SOD1 aggregate formation is not ordinarily observed in other forms of ALS, and the TDP-43 inclusions observed in the majority of other familial and sporadic ALS are not observed in SOD1 patients (Mackenzie et al., 2007), suggesting that an atypical molecular mechanism may be operating in these patients.

1.3.4 FTLD-tau

The microtubule associated protein tau ordinarily exists in a monomeric, soluble form within cells, and is thought to play a vital role in the assembly and stabilisation of microtubules (Spillantini and Goedert, 2013). Tau has been observed to localise to neuronal axons, and is thought to play a regulatory role in vesicular transport along microtubules (Spillantini and Goedert, 2013). In just under half of all FTLD cases, tau protein accumulates in an insoluble, hyperphosphorylated, filamentous form within neurons (Sieben et al., 2012). Aggregation of the tau repeat domain has been observed to be toxic to cultured cells (Khlistunova et al., 2006), however the mechanism by which this occurs has not been fully characterised. Mutations in the *MAPT* gene that encodes tau have also been observed to cause between 8 and 43 % of familial FTLD, depending on the population examined (Rademakers et al., 2004). Collectively, diseases involving tau neuropathology, including FTLD and Alzheimer's disease, are termed tauopathies. FTLD-tau disease subtypes include Pick's disease, corticobasal degeneration (CBD), progressive supranuclear palsy (PSP) and argyrophilic grain disease (AGD). A number of differentially spliced isoforms of the *MAPT* gene encoding tau exist, which can contain either three (3R) or four (4R) repeating microtubule-binding domains, and either zero (0N), one (1N) or two (2N) N-terminal inserts, which mediate interaction of tau protein with the plasma membrane. Insoluble tau in different tauopathies may have a differing composition of tau isoforms, and can differ between patients in FTLD (Spillantini and Goedert, 2013).

1.3.5 FTLD-UPS

Prior to the identification of TDP-43 and FUS as major inclusion components in some FTLD cases, all patients exhibiting ubiquitinated inclusions were classified in a single "FTLD-U" category. While most FTLD-U cases are now reclassified as FTLD-TDP or FTLD-FUS, a number of cases remain for which other components of the ubiquitinated inclusions have not been identified, including cases with familial FTD linked to mutations in the *CHMP2B* gene, which were observed to be negative for both TDP-43 and FUS (Holm et al., 2009, 2007). These cases are now re-classified as FTLD-UPS (Mackenzie et al., 2010a).

1.4 Genetics

Mutations within a number of genes have been found to cause familial forms of ALS, FTLD or both of these pathologies, summarised in Table 1.1 and Figure 1.2. As detailed in the previous section (1.3), mutations in all of the genes that encode the major aggregating proteins in ALS and FTLD, including *MAPT*, *TARDBP*, *FUS* and *SOD1* cause a subset of familial ALS and FTD cases. In addition, a number of mutations in other genes have also been found

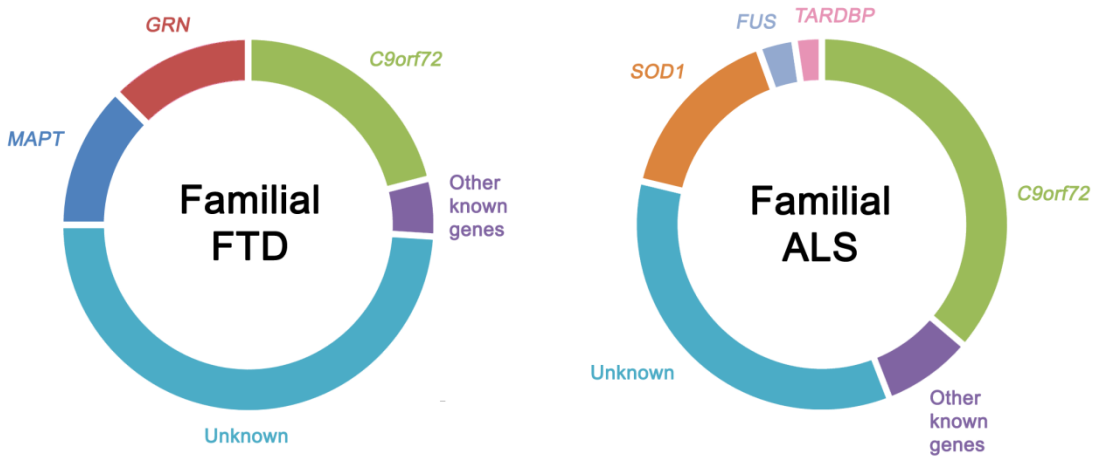


Figure 1.2 Genetic causes of familial FTD and familial ALS

Charts represent the approximate proportion of total familial FTD and familial ALS cases caused by mutations in individual genes. Familial FTD data was derived from: (Rademakers et al., 2012). Familial ALS data was derived from: (Renton et al., 2014).

to cause these pathologies, the most prevalent of which include *C9orf72* (which will be discussed in detail from section 1.5 onwards), and *GRN.*

GRN encodes progranulin, a growth factor precursor implicated in the promotion of neuronal growth and survival, as well as processes such as inflammation and wound repair (Ahmed et al., 2007). Mutations in *GRN* result in a loss of protein function, frequently by premature termination of the protein sequence (Gass et al., 2006), which causes between 3 and 26 % of all familial FTLD cases, depending on the population examined (Gijselinck et al., 2008). *GRN* mutations result in an FTLD-TDP type A pathology, and either bvFTD or PNFA (Sieben et al., 2012). However it is not yet fully understood how loss of progranulin function leads to the onset of FTLD.

The remaining genes are mutated in small numbers of ALS and FTLD patients, but share common themes that underline potentially important mechanisms within these disorders. Several genes play a role in autophagy and protein clearance, including *VCP*, *CHMP2B*, *UBQLN2*, *SQSTM1*, and *OPTN*, emphasising the importance of protein homeostasis and degradation in ALS and FTLD. A number of the mutated genes also play a role in RNA and/or DNA metabolism, including *HNRNPA1*, *TAF15*, *SPG11* and *SETX,* a function also played by aggregating proteins FUS and TDP-43, highlighting an important role for transcriptome changes in ALS and FTLD.

Gene symbol	Gene name	Protein function	Approx. frequency in familial ALS %	Approx. frequency in familial FTD %	Aggregating protein	Reference
FTD only						
MAPT	Microtubule associated protein tau	Cytoskeleton	-	8 - 43 %	Tau	(Hutton et al., 1998) (Rademakers et al., 2004)
GRN	Progranulin	Trophic factor Inflammation	-	3 – 26 %	TDP-43	(Cruts et al., 2006) (Baker et al., 2006) (Gijselinck et al., 2008)
Predominantly FTD						
VCP	Valosin-containing protein	Autophagy	Rare	Rare	TDP-43	(Watts et al., 2004) (Johnson et al., 2010)
CHMP2B	Charged multivesicular body protein 2B	Endosomal trafficking Autophagy	Rare	Rare	p62	(Skibinski et al., 2005) (Parkinson et al., 2006)
FTD / ALS						
C9orf72	Chromosome 9 open reading frame 72	Unknown (implicated in vesicular trafficking)	21 – 56 %	14 – 48 %	TDP-43	(DeJesus-Hernandez et al., 2011) (Renton et al., 2011a) (Majounie et al., 2012b)
UBQLN2	Ubiquilin 2	Autophagy	Rare	Rare	TDP-43, FUS	(Deng et al., 2011)
HNRNPA1	Heterogenous nuclear ribonucleoprotein A1	RNA metabolism	Rare	Rare		(Kim et al., 2013)
HNRNPA2B1	Heterogenous nuclear ribonucleoprotein A2/B1	RNA metabolism	Rare	Rare		(Kim et al., 2013)
SQSTM1	Sequestosome 1 (p62)	Autophagy	Rare	Rare		(Fecto, 2011) (Le Ber et al., 2013)
SIGMAR1	Sigma non-opioid intracellular receptor 1	Endoplasmic reticulum lipid rafts	Rare	Rare		(Luty et al., 2010)

Table 1.1 Details of genetic mutations found in FTD and ALS (continued overleaf)

Gene symbol	Gene name	Protein function	Approx. frequency in familial ALS %	Approx. frequency in familial FTD	Aggregating protein	Reference
Predominantly ALS						
FUS	Fused in sarcoma	DNA/RNA metabolism	4 %	Rare	FUS	(Vance et al., 2009) (Kwiatkowski et al., 2009) (Van Langenhove et al., 2010)
TARDBP	TAR DNA binding protein	DNA/RNA metabolism	3 %	Rare	TDP-43	(Neumann et al., 2006) (Lattante et al., 2013)
ALS only						
SOD1	Superoxide dismutase 1	Antioxidant	20 %	-	SOD1	(Rosen et al., 1993), (Pasinelli and Brown, 2006)
OPTN	Optineurin	Autophagy	Rare	-	TDP-43	(Maruyama et al., 2010)
ANG	Angiogenin	Blood vessel formation	Rare	-	TDP-43	(Greenway et al., 2006)
VAPB	VAMP (vesicle-associated membrane protein)-associated protein B	Vesicular trafficking	Rare	-	TDP-43	(Nishimura et al., 2004),
PFN1	Profillin 1	Actin dynamics	Rare	-		(Wu et al., 2012)
TAF15	TATA box binding protein (TBP)-associated factor 15	RNA metabolism	Rare	-		(Couthouis et al., 2011)
DAO	D-amino-acid oxidase	Peroxisomal enzyme	Rare	-		(Mitchell et al., 2010)
ALS2	Amyotrophic lateral sclerosis 2	Endosomal trafficking	Rare	-		(Hadano et al., 2001)
SPG11	Spastic paraplegia 11	DNA damage repair	Rare	-		(Daoud et al., 2012)
SETX	Senataxin	DNA/RNA metabolism	Rare	-		(Chen et al., 2004),
FIG4	Phosphoinositide 5-phosphatase	Lipid metabolism	Rare	-		(Chow et al., 2009)

Table 1.2 Details of genetic mutations found in FTD and ALS (continued from previous page)

1.5 The *C9orf72* repeat expansion mutation

1.5.1 Identification of a hexanucleotide repeat expansion mutation in *C9orf72*

In order to understand more about the genetic basis for the observed association between ALS and FTD, genome-wide linkage analysis studies were carried out by several groups on kindreds who exhibited concomitant ALS and FTD within single individuals or within different individuals in the same family. Collectively, these reports identified a minimum linkage region of 3.7 Mb on chromosome 9p21 conferring FTD/ALS susceptibility to individuals within these families (Boxer et al., 2011; Gijselinck et al., 2010; Le Ber et al., 2009; Luty et al., 2008; Morita et al., 2006; Pearson et al., 2011; Valdmanis et al., 2007; Vance et al., 2006). Concurrently, several independent genome-wide association studies (GWAS) performed on both ALS patients (Laaksovirta et al., 2010; Shatunov et al., 2010; van Es et al., 2009) and FTD patients (Van Deerlin et al., 2010) highlighted a risk haplotype in the same linkage region and narrowed the area of interest to a 232 Kb linkage disequilibrium block containing three genes: *MOBKL2B, INFK* and *C9orf72*. Initial sequencing of the protein coding regions of these genes found no evidence for exonic mutations linked to the pathology (Pearson et al., 2011), implicating non-coding DNA within this region in FTD/ALS susceptibility. Several large-scale international collaborations finally resulted in the identification of a GGGGCC hexanucleotide repeat expansion within the first intron of *C9orf72* (DeJesus-Hernandez et al., 2011; Gijselinck et al., 2012; Renton et al., 2011a), which initially evaded detection in part due to the difficulty in sequencing this repetitive, GC-rich sequence.

1.5.2 Genetic characteristics of the *C9orf72* repeat expansion

Three transcript variants are generated from the *C9orf72* gene: variants 1 and 3 encode a 481 amino acid long protein isoform C9orf72a, while variant 2 encodes a shorter 222 amino acid isoform C9orf72b (Figure 1.3a). The repeat expansion is present within the promotor

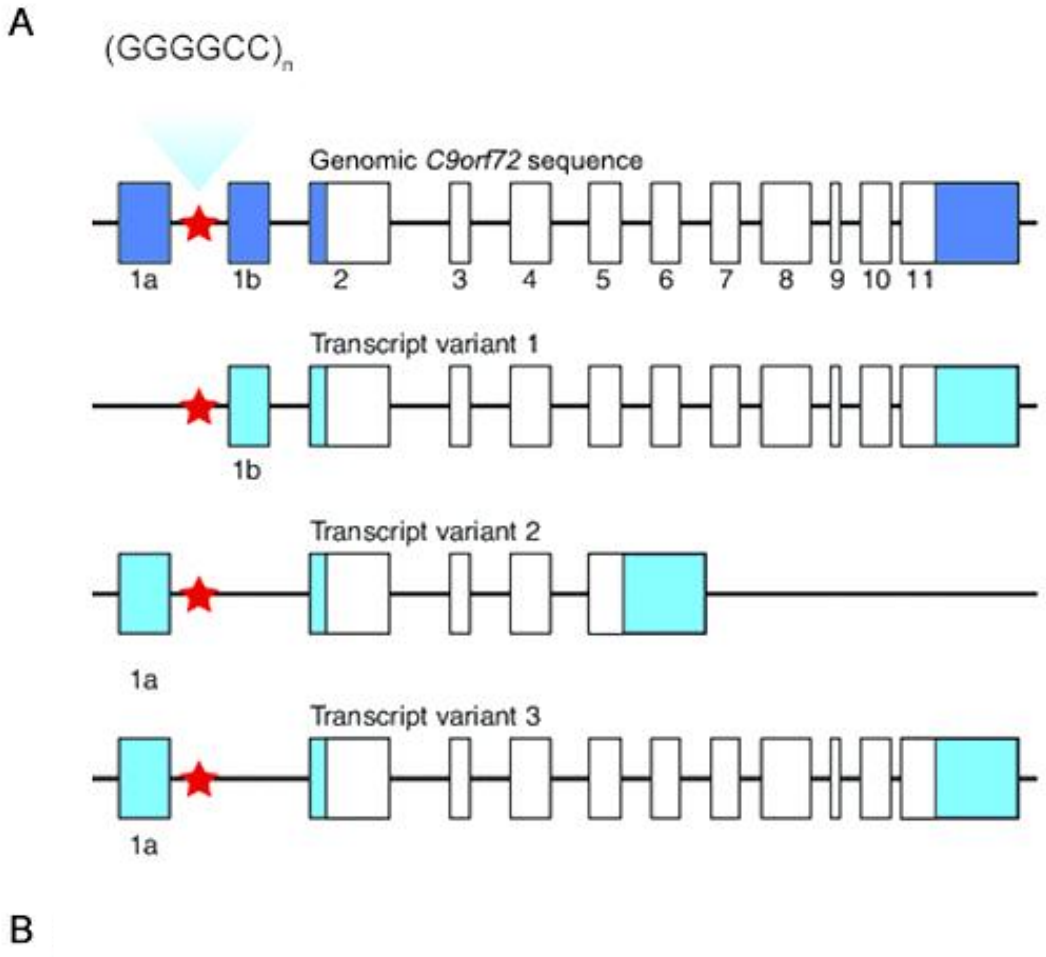


Figure 1.3 The *C9orf72* repeat expansion

A) The *C9orf72* gene locus and transcript variants. Red star represents the locus of the GGGGCC repeat expansion. Numbered rectangles represent coding (white) and non-coding (blue) exons in the *C9orf72* gene. Adapted from (Woollacott and Mead, 2014). B) Example electropherograms of repeat-primed PCR reactions of expanded and normal *C9orf72* repeats. Adapted from (Rademakers et al., 2012).

region of variant 1, but falls within the first intron of variants 2 and 3, between non-coding exons 1a and 1b, permitting the transcription of the repeat expansion itself into RNA in these transcripts.

Normal individuals exhibit a non-expanded repeat size of between 2 and around 30 GGGGCC units in *C9orf72*, with two repeats most frequently observed (DeJesus-Hernandez et al., 2011; Ratti et al., 2012), while repeat expansion carriers most often exhibit between 800 and 4400 repeats (Beck et al., 2013). However, the lower limit of pathogenic repeat lengths has not been definitively determined, with some mild cognitive dysfunction reported in patients with intermediate length repeats (Byrne et al., 2014; van der Zee et al., 2013). Expanded repeats are also subject to extreme somatic instability and differ in length between different patient tissues, and from generation to generation, with some recent evidence suggesting increased repeat length and disease anticipation in parent to child transmission (Gijselinck et al., 2015).

Due to the repetitive and GC-rich nature of the repeat expansion, sequencing an expansion comprised of thousands of thousands of repeats in its entirety is unfeasible. Presence of the repeat expansion is therefore commonly determined using repeat-primed PCR, which uses three primers: a forward and a reverse primer targeted to the sequences flanking the repeats, plus an additional reverse primer complimentary to the repeat sequence itself. In a large GGGGCC repeat expansion, the repeat primer can bind at numerous positions, and so yields an array of different length amplification products that diminish in concentration with increasing length to give a characteristic saw-tooth amplification pattern indicating the presence of a repeat (Figure 1.3b). However, repeat sequences can only be successfully amplified up to a maximum of approximately 60 repeats, meaning that while the presence or absence of a large expansion can be determined, other techniques are necessary to determine its length. To achieve this, Southern blotting can be carried out to approximate

the repeat size or range of repeat sizes present in patient tissue. This can be accomplished by rigorous restriction enzyme digestion of genomic patient DNA, followed by electrophoretic resolution and transfer of the digested DNA and subsequent hybridisation with a labelled DNA probe complimentary to the repeat expansion (Beck et al., 2013).

1.5.3 Epidemiology of *C9orf72* FTD/ALS

The hexanucleotide repeat expansion in *C9orf72* is presently the most prevalent known cause of ALS, FTD, or combined FTD/ALS, exceeding the incidence of any other pathological mutation. The expansion is most frequently found in European and North American Caucasian populations, with a particularly high frequency in certain Nordic populations such as Finland, Sweden and Denmark, with 28 % of FTD and 29 % of ALS patients in Finland found to be positive for the expansion (Renton et al., 2011a; van der Zee et al., 2013). In the UK, 11 % of FTD and 8 % of ALS patients have been found to be positive for the expansion (Majounie et al., 2012b). Conversely, the expansion is rare within East Asian populations, with the exception of infrequent hubs of high prevalence such as in the Kii peninsula in Japan (Ishiura et al., 2012). Information about prevalence in South Asian, South American, African and Middle Eastern populations is currently lacking (Woollacott and Mead, 2014). The high prevalence of the repeat expansion in specific Northern European populations combined with the fact that patients with the repeat expansion share a common risk haplotype on adjacent regions of chromosome 9 has led some groups to hypothesise that the mutation arose from a single expansion event in a European ancestor six thousand years ago (Smith et al., 2013). However, other studies have disputed the existence of a single common ancestor, instead suggesting that *de novo* expansion of the repeats is more favourable on the risk haplotype, resulting in multiple expansion events in unrelated individuals throughout history (Beck et al., 2013; Dobson-Stone et al., 2013)

1.5.4 Clinical features of *C9orf72* FTD/ALS

A large degree of clinical diversity is observed in individuals bearing the *C9orf72* repeat expansion, and patients can develop ALS, FTD or combined FTD/ALS. Patient series with mixed FTD/ALS disease course exhibit a notably high *C9orf72* mutation frequency of up to 88 %, depending on the cohort (Cruts et al., 2013). While the *C9orf72* repeat expansion is present in a higher proportion of familial cases of ALS and FTD, it is also found in some sporadic cases, which is of particular importance in ALS wherein 90 % of all cases are in patients lacking a family history of the disease. These non-familial cases may be due to *de novo* expansion of the unstable repeats between generations, however many apparently sporadic cases may be explained by the fact that the pathology does not manifest in all individuals with the mutation, and therefore the presence of the expansion in previous generations can be masked. Disease penetrance in patients with the *C9orf72* repeat expansion is almost complete, with 50 % penetrance in ALS and FTD patients at 58 years of age, and almost full penetrance by 80 years of age (Majounie et al., 2012b). However, disease onset has been observed in patients in their ninth decade, and healthy elderly individuals have been found to carry the expansion (Majounie et al., 2012b). Conversely, disease onset has also been observed in patients as young as 27 years old (Majounie et al., 2012b). Disease duration is heavily dependent on the manifestation of disease subtype in each patient. *C9orf72* ALS patients were found to have an average duration of 33 ± 2 months (mean ± standard error) by one study (Millecamps et al., 2012), while *C9orf72* FTD patients characteristically experience a much longer disease course similar to other inherited forms of FTD, with an average duration of 74 ± 11 months (mean ± standard error) observed by another report (Van Langenhove et al., 2013).

Motor neuron disease exhibited by patients with the *C9orf72* repeat expansion is almost exclusively ALS, with rare incidences of PMA, PLS or PBP (Snowden et al., 2012; Stewart et al., 2012). While most *C9orf72* ALS cases exhibit limb-onset, a higher proportion of bulbar-

onset is observed in patients with the repeat expansion compared to those without (Majounie et al., 2012b). Minor cognitive dysfunction not sufficiently severe to be classed as FTD is also more common in *C9orf72* ALS patients compared to those without the repeat expansion (Byrne et al., 2012).

C9orf72 FTD cases present with bvFTD in 87 % of cases, with a minority of patients presenting with PNFA and rare incidences of SD reported (Boeve et al., 2012; Van Langenhove et al., 2012). Psychiatric symptoms such as anxiety, agitation and psychosis are more frequently observed in *C9orf72* FTD patients compared to patients without the repeat expansion (Mahoney et al., 2012; Snowden et al., 2012), and a proportion of *C9orf72* FTD patients additionally present with some Parkinsonism, with an akinetic-rigid syndrome or tremor observed (Boeve et al., 2012).

Rarely, *C9orf72* repeat expansions have been observed in patients with Parkinson's disease and Alzheimer's disease: in a study examining a large Parkinson's Disease cohort, 0.055 % of patients were found to carry the expansion, all of whom presented with atypical Parkinson's Disease with elements of cognitive decline (Theuns et al., 2014). In a cohort of patients diagnosed with Alzheimer's Disease, less than 1 % of patients carried the *C9orf72* repeat expansion, and several of the patients presented with ubiquitin-positive inclusions characteristic of FTD when subsequent post-mortem examination was carried out (Majounie et al., 2012a). These results suggest that Parkinson's disease and Alzheimer's disease do not represent major manifestations of the *C9orf72* repeat expansion, and some of the reported cases may be attributable to misdiagnosis of FTD presenting with atypical symptoms.

1.5.5 Neuroanatomical features of *C9orf72* FTD/ALS

Neuroimaging studies have revealed some characteristic differences in the distribution of brain atrophy in patients with the *C9orf72* repeat expansion compared to patients with

other forms of FTD and ALS. Magnetic Resonance Imaging (MRI) studies of FTD patients found that *C9orf72* FTD patients were associated with a symmetric pattern of frontal, temporal and parietal cortical atrophy in contrast to the asymmetric atrophy exhibited by patients with *GRN* mutations (Mahoney et al., 2012; Whitwell et al., 2012). *C9orf72* FTD patients also exhibited an unusual degree of cerebellar and thalamic atrophy compared to other FTD subtypes, with less temporal lobe involvement observed compared to *MAPT* mutation bearing patients (Mahoney et al., 2012; Whitwell et al., 2012). ALS patients with the *C9orf72* repeat expansion were also observed to exhibit a distinct pattern of atrophy compared to other ALS patients, with *C9orf72* ALS patients exhibiting a greater degree of extramotor abnormality within frontal cortical, temporal cortical and thalamic regions, in accord with the higher incidence of cognitive dysfunction observed in these patients compared to other types of ALS (Bede et al., 2013; Byrne et al., 2012).

1.5.6 Neuropathology in *C9orf72* FTD/ALS

Cases with the *C9orf72* repeat expansion typically exhibit TDP-43 pathology, including neuronal and glial cytoplasmic inclusions, diffuse cytoplasmic TDP-43, dystrophic neurites and infrequent neuronal intranuclear inclusions (Mackenzie et al., 2014). In patients with pure ALS, prominent TDP-43 pathology is found in upper motor neurons, brainstem and spinal cord, while patients with pure FTLD show abundant TDP-43 pathology in the frontal and temporal cortex, and patients with combined FTLD-ALS exhibit a degree of pathology in both of these areas. A good correlation is usually observed between TDP-43 inclusion burden and the degree of degeneration in each brain region (Mackenzie et al., 2013).

In cases that exhibit some degree of cortical TDP-43 pathology, an FTLD-TDP type B pathology is most often observed, with neuronal cytoplasmic inclusions present throughout the cortex and infrequent neuronal intranuclear inclusions and dystrophic neurites. A smaller subset of patients also present with FTLD-TDP type A pathology, which is typified by

small neuronal cytoplasmic inclusions and frequent, short dystrophic neurites predominantly in layer II of the neocortex, with some neuronal intranuclear inclusions. FTLD-TDP type A pathology is more frequently found in patients with a pure FTLD pathology, while type B is frequently associated with a mixed FTLD-ALS pathology. Rare *C9orf72* repeat expansion cases have been found that lack TDP-43 pathology: one *C9orf72* FTLD case from a Belgian cohort was found to be absent for TDP-43 inclusions (Gijselinck et al., 2012), while another patient in the UK exhibited FTLD-tau pathology with no TDP-43 inclusions (Snowden et al., 2012). Additionally, several *C9orf72* repeat expansion bearing patients who died prematurely have been studied, and have been observed to exhibit very sparse or no TDP-43 pathology (Baborie et al., 2014; Proudfoot et al., 2014)

In addition to TDP-43 inclusions, *C9orf72* FTD/ALS patients also exhibit characteristic TDP-43 negative, p62 positive inclusions, which can take the form of "star-like" cytoplasmic inclusions or "dot-like" intranuclear inclusions (Al-Sarraj et al., 2011). These inclusions are present in most brain regions and in the spinal cord, with particularly high abundance throughout the cortex, cerebellum and hippocampus (Schludi et al., 2015). In particular, cerebellar p62 positive inclusions are a defining hallmark of *C9orf72* pathology that are rarely observed in any other FTD or ALS disease subtypes (Al-Sarraj et al., 2011). In 2013, it was found that these inclusions also contain five different proteins consisting of repeating dipeptide units, translated from the repeat expansion itself by a novel process called repeat-associated, non-ATG initiated (RAN) translation, which will be discussed in more detail in section 1.6.3. A subset of these aggregates have additionally been found to contain the RNA binding protein hnRNP A3 (Mori et al., 2013b).

C9orf72 FTD/ALS patient neurons also exhibit aggregates of RNA termed RNA foci comprised of sense GGGGCC and antisense CCCCGG RNA repeats transcribed from the expansion (Cooper-Knock et al., 2015b, 2014; Donnelly et al., 2013; Gendron et al., 2013;

Lagier-Tourenne et al., 2013; Lee et al., 2013; Mizielinska et al., 2013; Zu et al., 2013), which accumulate within the nucleus and to a lesser extent in the cytoplasm (Mizielinska et al., 2013). The role of RNA foci in *C9orf72* FTD/ALS will be discussed more in section 1.6.2.

1.6 Potential disease mechanisms in *C9orf72* FTD/ALS

Three main mechanisms of toxicity have been implicated in *C9orf72* FTD/ALS pathology (Figure1.4): 1) a loss of C9orf72 protein function due to disruption of gene transcription by the repeats (section 1.6.1), 2) a gain-of-RNA function as a result of transcription of the repeat DNA into repetitive RNA in the sense and antisense orientations, which subsequently aggregate within the cell to form RNA foci (section 1.6.2) and 3) a gain-of-

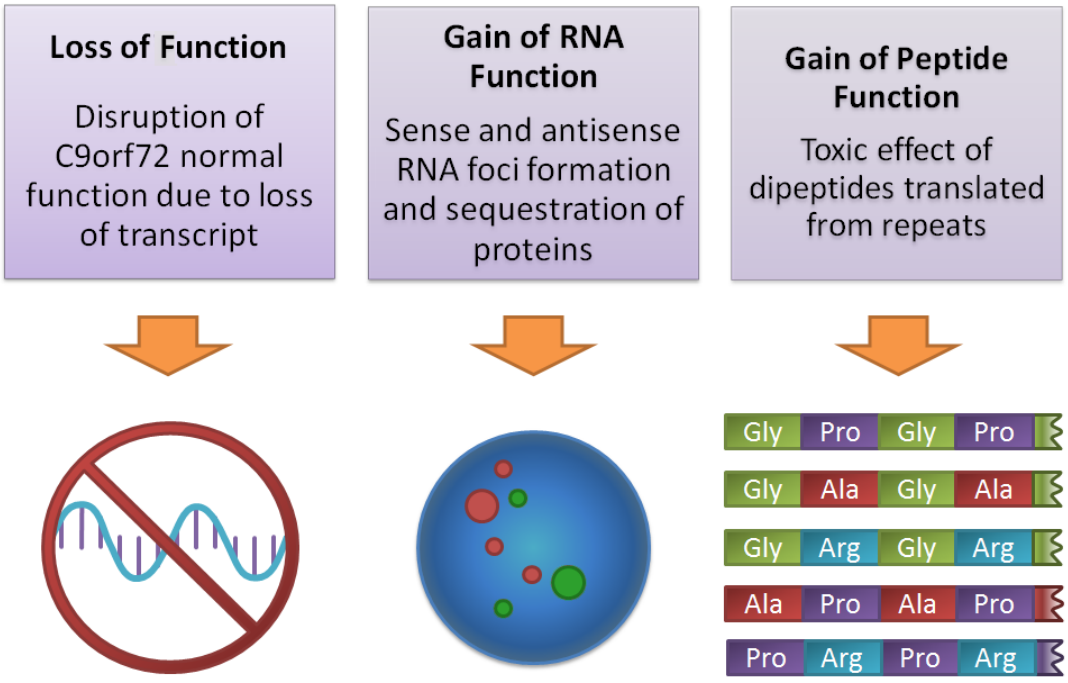


Figure 1.4 Potential disease mechanisms in *C9orf72* FTD/ALS.

Illustration summarises the three main loss and gain-of-function mechanisms that may contribute to pathogenesis in *C9orf72* FTD/ALS. Loss-of-function may arise from a reduction of *C9orf72* transcripts, depicted here by an RNA molecule under a cancel sign, leading to a potential reduction in normal C9orf72 protein function. Gain of RNA function may occur as a result of the sequestration of vital RNA binding proteins into RNA foci, depicted here by sense (red circles) and antisense (green circles) foci within a cell nucleus (blue circle) as seen in fluorescence *in situ* hybridisation experiments. Gain of peptide function may arise from the translation of sense and antisense RNA repeat transcripts into five different repeating dipeptides, with the proteins produced depicted here using their respective three letter amino acid codes: gly = glycine, pro = proline, ala = alanine, arg = arginine.

peptide function due to the translation of repeat RNA into proteins in all sense and antisense frames, forming five different dipeptide repeat proteins that aggregate within neurons (section 1.6.3).

1.6.1 Loss of C9orf72 protein function

The loss of a protein resulting in the disruption of a vital biological process is a common disease mechanism, and is observed in a number of repeat expansion disorders including fragile X syndrome, in which a CGG repeat expansion in the 5' untranslated region of *FMR1* causes a loss of fragile X mental retardation protein (FMRP)(Penagarikano et al., 2007), and Friedreich's ataxia, in which a GAA repeat expansion in the first intron of *FXN* causes a loss of frataxin protein (Delatycki et al., 2000). A number of studies have therefore investigated whether a loss of normal C9orf72 protein function could contribute to *C9orf72* FTD/ALS.

While the exact function of C9orf72 protein remains to be determined, it is known to share homology with the differentially expressed in normal and neoplasia (DENN) family of GDP/GTP exchange factors (GEFs) which activate Rab-GTPases and modulate vesicular trafficking (Levine et al., 2013; Zhang et al., 2012). In support of this function, siRNA knockdown of C9orf72 protein was found to disrupt transport of shiga toxin, the receptor TrkB and autophagosome marker light chain 3 (LC3), implicating C9orf72 protein in the modulation of autophagy and endocytosis (Farg et al., 2014). C9orf72 protein was also found to colocalise with Rab proteins in neuronal cell lines, primary rat cortical neurons and in human post-mortem spinal cord sections (Farg et al., 2014). However, study of C9orf72 localisation is presently limited by the currently available antibodies, which exhibit non-specific staining. Some of the reports published using commercially available anti-C9orf72 antibodies give divergent observations of C9orf72 localisation, and lack negative controls (Cooper-Knock et al., 2012; DeJesus-Hernandez et al., 2011; Farg et al., 2014; Gijselinck et al., 2012; Hsiung et al., 2012; Snowden et al., 2012; Stewart et al., 2012). New antibodies

were more recently generated that are reported to differentially recognise the two C9orf72 protein isoforms. Using these antibodies, the long C9orf72 isoform was observed to exhibit diffuse cytoplasmic localisation, meanwhile the short C9orf72 isoform was found to localise to the nuclear membrane in neurons from non-neurodegenerative control patients, suggesting a differential localisation of the two isoforms. Both long and short C9orf72 isoforms were reported to interact with components of the nuclear pore complex including Ran-GTPase and Importin β1 (Xiao et al., 2015), suggesting a possible function in nuclear transport. However, in motor neurons of ALS patients both with and without the repeat expansion, the short isoform was not observed at the nuclear membrane and instead localised to the plasma membrane (Xiao et al., 2015), suggesting a disruption in short-isoform C9orf72 protein localisation in disease. However further antibody characterisation and staining of C9orf72 protein deficient tissue as a negative control is needed to confirm these observations.

Several studies have observed a significant reduction in total *C9orf72* RNA (Belzil et al., 2013; Fratta et al., 2013; van Blitterswijk et al., 2015; Waite et al., 2014), as well as a specific reduction in transcript variant 1 (Belzil et al., 2013; DeJesus-Hernandez et al., 2011; Fratta et al., 2013; van Blitterswijk et al., 2015), the promotor region of which is disrupted by the presence of the repeat expansion. One report additionally observed a specific reduction in transcript variant 2 (van Blitterswijk et al., 2015), but a significant reduction in variant 3 has not been observed (Belzil et al., 2013; DeJesus-Hernandez et al., 2011; Fratta et al., 2013; van Blitterswijk et al., 2015; Waite et al., 2014). A decrease in C9orf72 protein level has also been reported in *C9orf72* FTD post-mortem frontal cortex (Waite et al., 2014). A study examining an FTD patient homozygous for the *C9orf72* repeat expansion additionally found that the presence of two expanded *C9orf72* repeat alleles in this patient is associated with a decrease in the levels of variant 1 and variant 2 RNA transcripts compared to heterozygous *C9orf72* expansion patients and control patients, however

transcripts of all variants were still present within the homozygous patient, suggesting an incomplete loss of C9orf72 protein even in the presence of double repeat expansions (Fratta et al., 2013).

The reduction in *C9orf72* transcription is thought to be due to epigenetic silencing. Methylation of cytosine-phosphate-guanine (CpG) dinucleotides is an important mechanism in inhibiting gene expression, and is associated with the induction of a repressed chromatin state. Hypermethylation of CpG sites adjacent to a repeat expansion has been observed in several other disorders including Friedreich's ataxia (Al-Mahdawi et al., 2008; Evans-Galea et al., 2012; Greene et al., 2007), fragile X syndrome (Bell et al., 1991; Pieretti et al., 1991; Sutcliffe et al., 1992) and myotonic dystrophy (Klesert et al., 1997; Korade-Mirnics, 1999; Thornton et al., 1997). Two CpG islands are present upstream of the *C9orf72* repeat expansion, and were found to be hypermethylated in up to 36 % of repeat expansion carriers, with a similar level of methylation seen in both ALS and FTD patients (Liu et al., 2014; Russ et al., 2015; Xi et al., 2014, 2013). Furthermore, CpG sequences within the repeat itself are also methylated in 97 % of patients with over 50 repeats (Xi et al., 2015b). The degree of methylation in the *C9orf72* promotor was found to be associated with a reduction in *C9orf72* RNA transcript levels in patient derived lymphoblast lines but was reversed following treatment with the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine (Liu et al., 2014). Trimethylated histone residues, which are also known to inhibit gene expression, were additionally found to bind to the *C9orf72* repeat expansion in patient derived fibroblast lines (Belzil et al., 2013).

A number of studies have been published that support a role for loss of C9orf72 protein function in *C9orf72* FTD/ALS. Knockdown of the zebrafish *C9orf72* homolog (*zC9orf72*) using antisense morpholino oligonucleotides to block translation of *zC9orf72* RNA transcripts resulted in axonal degeneration in motor neurons and defective locomotion in

swimming tasks (Ciura et al., 2013). Deletion of *Caenorhabditis elegans C9orf72* homolog *alfa-1* resulted in a progressive paralysis phenotype in the worms and the degeneration of motor neurons (Therrien et al., 2013). Furthermore, one study found that patients with the lowest 25th percentile of *C9orf72* variant 1 expression had a significantly shorter disease duration than the top 75th percentile of patients, indicating a more aggressive disease course in patients with low *C9orf72* transcripts (van Blitterswijk et al., 2013), while the degree of *C9orf72* promoter methylation also correlated with a lower age of onset and a higher repeat length (Gijselinck et al., 2015).

However, other evidence suggests that loss of C9orf72 protein function does not represent the primary pathological mechanism in *C9orf72* FTD/ALS. Firstly, no cases of ALS or FTD have been found that are caused by loss-of-function mutations in the coding region of *C9orf72* (Harms et al., 2013), suggesting that the repeat expansion itself is important in pathogenesis. Knockdown of *C9orf72* was additionally found not to be toxic in *C9orf72* ALS patient derived iPSCs (Sareen et al., 2013). Several C9orf72 loss-of-function mouse models have also failed to induce a neurodegenerative phenotype: mice treated with antisense oligonucleotides against *C9orf72* did not exhibit any behavioural or pathological neurodegenerative features (Lagier-Tourenne et al., 2013), and conditional knock-out of *C9orf72* in neurons and glia of *Nestin-Cre$^{+/-}$;C9orf72$^{fl/fl}$* mice resulted in decreased body weight, but no neurodegeneration, motor defects or reduction in viability (Koppers et al., 2015).

Silencing of the *C9orf72* locus may even serve a protective role in patients due to the attenuation of downstream gain-of-function disease mechanisms. Hypermethylation of the *C9orf72* promoter in post-mortem FTD/ALS patient brain was associated with a reduction in the percentage of cells exhibiting RNA foci or DPR protein aggregates (Liu et al., 2014), and was additionally associated with a reduction in the loss of grey matter density (McMillan et

al., 2015). One study also observed that *C9orf72* promotor methylation in the brain and blood is correlated with a later age of death, with methylation in the blood found to be associated with a longer disease duration in FTD patients (Russ et al., 2015).

These data therefore suggest that while loss of C9orf72 protein function may exacerbate disease progression, it does not represent the major pathological pathway, and reduction of harmful repeat RNA and DPR proteins through silencing of the *C9orf72* gene locus at the cost of a reduction in C9orf72 protein levels may therefore still exhibit a beneficial net effect.

1.6.2 RNA gain-of-function mechanisms

An RNA gain-of-function disease mechanism has been proposed in a number of non-coding repeat expansion diseases and is best characterised in myotonic dystrophy type 1 (DM1), in which a CTG repeat in the 3' untranslated region (UTR) of the human *DMPK* gene is expanded from 5-38 repeats in control patients, to between 50 and several thousands of repeats in length (Brook et al., 1992; Fu et al., 1992; Mahadevan et al., 1992). Following transcription into RNA, the CUG repeats form a hairpin secondary structure, and aggregate to form RNA foci in the cell nucleus (Koch and Leffert, 1998; Taneja, 1995). These foci sequester a number of RNA-binding proteins including the splicing protein muscleblind-like protein 1 (MBNL1), leading to a loss of normal splicing function (Mankodi, 2001; Miller et al., 2000). This results in the aberrant splicing of chloride channel ClC-1, a vital component in normal muscle function, and the expression of embryonic as opposed to the adult splice variant in adult muscles (Mankodi et al., 2002). This dysregulated expression gives rise to the hyperexcitability of muscle fibre membranes, resulting in the symptoms of myotonia (Mankodi et al., 2002). Therefore, a clear molecular cascade can be observed leading from the expression of the repeats to the disease phenotype seen in patients.

A similar mechanism has been proposed to operate in *C9orf72* FTD/ALS. Studies using RNA fluorescence in-situ hybridisation have shown that both sense (GGGGCC) and antisense (CCCCGG) repeat transcripts form RNA foci in human brain (Cooper-Knock et al., 2015b, 2014; Donnelly et al., 2013; Gendron et al., 2013; Lagier-Tourenne et al., 2013; Lee et al., 2013; Mizielinska et al., 2013; Zu et al., 2013) (Figure 1.5a). Sense and antisense foci have been observed in neurons throughout the CNS including the frontal cortex, motor cortex, cerebellum, spinal cord and hippocampus (Cooper-Knock et al., 2015b, 2014; DeJesus-Hernandez et al., 2011; Donnelly et al., 2013; Gendron et al., 2013; Lagier-Tourenne et al., 2013; Lee et al., 2013; Mizielinska et al., 2013; Zu et al., 2013), and are primarily localised within the cell nucleus, with a minority found in the cytoplasm (Lagier-Tourenne et al., 2013; Mizielinska et al., 2013). RNA foci are additionally found in glial cells, but to a lesser extent (Gendron et al., 2013; Lagier-Tourenne et al., 2013; Mizielinska et al., 2013) and have also been detected in cell cultures derived from tissues outside of the CNS such as fibroblasts and lymphoblasts (Cooper-Knock et al., 2014; Zu et al., 2013). Sense RNA foci formation has been observed in a mouse model expressing 66 GGGGCC repeats in the CNS (Chew et al., 2015), while both sense and antisense foci formation have been observed in *C9orf72* patient derived iPSCs, and neurons differentiated from these iPSCs (Almeida et al., 2013; Donnelly et al., 2013; Sareen et al., 2013), as well as in a number of transgenic cell lines overexpressing the repeats (Gendron et al., 2013; Lee et al., 2013; Mizielinska et al., 2014; Rossi et al., 2015)

In *C9orf72* FTD patients, RNA foci were observed to be most abundant in the frontal cortex, the brain region most affected in FTD, with an inverse correlation between the frequency of cortical sense foci and the age of disease onset (Mizielinska et al., 2013), while the presence of antisense foci was observed to correlate with TDP-43 mislocalisation in C9-ALS motor neurons (Cooper-Knock et al., 2015b), consistent with a role for RNA mediated toxicity.

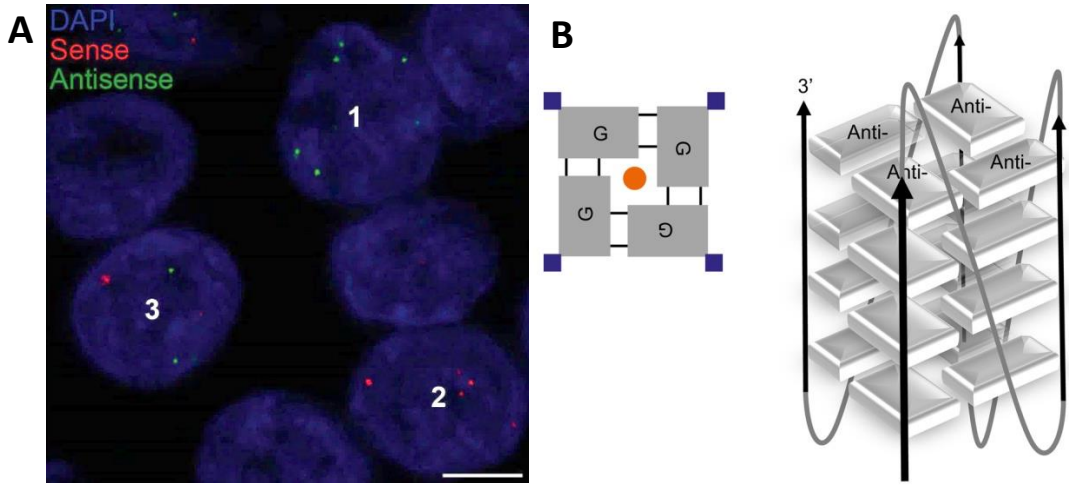


Figure 1.5 RNA foci and G-quadruplex formation

A) Confocal fluorescence microscopy image taken of RNA fluorescence in-situ hybridisation in *C9orf72* FTD patient frontal cortex, showing DAPI stained nuclei (blue), sense RNA foci (red) and antisense RNA foci (green). Cell 1 exhibits only antisense foci, cell 2 exhibits only sense foci, and cell 3 exhibits a mixture of both sense and antisense foci. Figure taken from (Mizielinska et al., 2013). B) Schematic of G-quadruplex secondary structure formed by *C9orf72* repeat RNA. Left hand image shows a single G quartet comprised of guanine residues (grey) stabilised by hydrogen bonds (black) around a central metal ion (orange), with surrounding phosphate backbone (blue). Right hand image shows four stacked G-quartets with anti glycosidic bonds in a parallel conformation. Figure taken from (Fratta et al., 2012).

GGGGCC repeat RNA and DNA form a stable secondary structure common to repetitive guanine rich sequences called a G-quadruplex (Fratta et al., 2012; Haeusler et al., 2014; Reddy et al., 2013). Sets of four guanine nucleotides form hydrogen bonds in a planar structure around a monovalent cation to form G-quartets, which stack together to form a G-quadruplex (Bochman et al., 2012)(Figure 1.5b). G-quadruplexes have been implicated in a range of different biological processes, such as transcriptional regulation, maintenance of telomeres, splicing and RNA transport (Millevoi et al., 2012). GGGGCC RNA repeats were shown to form a parallel G-quadruplex structure (Fratta et al., 2012; Reddy et al., 2013), while DNA repeats predominantly form an anti-parallel structure (Haeusler et al., 2014). These structures can be either unimolecular, with one RNA or DNA molecule folding upon itself, or multimolecular, in which several RNA and DNA molecules form the structure together (Reddy et al., 2013). It has been observed that the formation of G-quadruplexes in GGGGCC repeat DNA contributes to decreased RNA polymerase processivity (Haeusler et al., 2014). Additionally, repeat RNA was found to associate with repeat DNA after

transcription, forming a structure termed an R-loop, which further contributes to the abortion of transcripts (Haeusler et al., 2014). When transfected into cells, these truncated transcripts were observed to reduce cell viability (Haeusler et al., 2014), suggesting that G-quadruplexes and other secondary nucleic acid structures could play an important role in disease pathology. Furthermore, the prolonged interaction of RNA and DNA repeats in R-loops is hypothesised to contribute to DNA slippage and repeat instability due to the prolonged disassociation of the sense and antisense DNA strands (Haeusler et al., 2014).

A number of proteins have been proposed to recognise and interact with sense repeat RNA, antisense repeat RNA and the G-quadruplex motif, and may therefore be sequestered by the repeats. Common themes in the function of these interacting proteins include splicing, RNA trafficking, and nuclear import and export. Several heterogeneous nuclear ribonucleoproteins (hnRNPs), a class of proteins abundant in the nucleus with diverse RNA processing roles (Krecic and Swanson, 1999), have been observed to co-localise with both sense and antisense foci in human patient brain; these include hnRNP A1, hnRNP H and hnRNP F (Cooper-Knock et al., 2015b, 2014; Donnelly et al., 2013; Lee et al., 2013), factors which are known to play an important role in mRNA splicing (Krecic and Swanson, 1999). However not all studies were able to observe the sequestration of these proteins, with the absence of hnRNP A1 sequestration noted in sense RNA foci in *C9orf72* ALS patient cerebellum tissue (Donnelly et al., 2013), and no sequestration of hnRNP H or hnRNP F observed in a study in *C9orf72* FTD patient derived iPSCs (Almeida et al., 2013). hnRNP K, a protein known to bind cytosine-rich sequences with multiple functions in splicing, transcription and translation (Bomsztyk et al., 2004), was shown to bind preferentially to GGCCCC antisense repeat RNA (Haeusler et al., 2014), and co-localised with antisense RNA foci in *C9orf72* ALS patient cerebellum (Cooper-Knock et al., 2015b). However a crosslinking assay failed to show a direct interaction between hnRNP K and repeats in patient tissue (Cooper-Knock et al., 2015b). Additionally, hnRNP A3, which plays a role in cytoplasmic RNA

trafficking (Ma et al., 2002), was also observed to bind GGGGCC repeats *in vitro* (Mori et al., 2013b) but was not observed to co-localise with sense RNA foci in *C9orf72* FTD/ALS patient cerebellum (Lee et al., 2013). The degree to which sequestration of hnRNPs impacts on splicing and other RNA processing activities, is therefore yet to be determined.

A number of other RNA processing factors have also been observed to interact with repeat RNA under certain conditions. The mRNA nuclear export factor Aly/REF, which belongs to a family of hnRNP-like proteins (Stutz et al., 2000), was observed to co-localise with both sense and antisense RNA foci in *C9orf72* ALS patient brain (Cooper-Knock et al., 2015b, 2014), as was serine/arginine-rich splicing factor 2 (SRSF2), a component of the spliceosome, (Cooper-Knock et al., 2015b, 2014; Lee et al., 2013). Adenosine deaminase RNA specific B2 (ADARB2), one of a family of ADAR RNA editing proteins that are enriched in the CNS, was also observed to co-localise with sense RNA foci (Donnelly et al., 2013). The ADAR proteins convert adenosine to inosine, which is an important step in the production of the AMPA receptor subunit gluA2, and the downregulation of this protein has been implicated in sporadic ALS (Hideyama et al., 2012). However, ADARB2 is thought to lack this activity, although it has been hypothesised to play a complimentary role (Chen et al., 2000). Pur-alpha, a protein known to bind and transport purine rich RNA and DNA, was observed to interact with GGGGCC repeats in a pull down assay from human brain lysate (Xu et al., 2013), and co-localised with sense RNA foci in neurons derived from *C9orf72* patient iPSCs (Sareen et al., 2013). It has also previously been shown that pur-alpha binds to the CGG repeats found in fragile X tremor/ataxia syndrome (FXTAS) in a CCG repeat expressing fly model as well as CCG RNA foci in FXTAS patient brain (Jin et al., 2007). However, several other studies were unable to replicate the interaction of pur-alpha with GGGGCC repeat RNA foci in *C9orf72* ALS patient cerebellum or in *C9orf72* ALS patient derived iPSCs (Donnelly et al., 2013; Lee et al., 2013). It is therefore currently unclear to what degree

sequestration of RNA binding proteins plays a role in *C9orf72* RNA gain-of-function pathology.

1.6.3 Gain of peptide function

In 2011 it was discovered that the non-coding CAG repeat expansions in myotonic dystrophy and spinocerebellar ataxia type 8 (SCA8) are translated in the absence of an ATG methionine start codon by an non-canonical process called repeat associated non-ATG initiated (RAN) translation (Zu et al., 2011). This translation occurs in all reading frames, yielding three different homopolymeric proteins: polyglutamine, polyalanine and polyserine (Zu et al., 2011). Expression of these homopolymeric proteins was detected in CAG repeat expressing cell models, in mouse models of SCA8 and myotonic dystrophy, and in myoblasts from myotonic dystrophy patients. Furthermore, polyglutamine aggregates in the myotonic dystrophy mouse model were associated with markers of apoptosis, suggesting that these proteins are relevant to disease progression (Zu et al., 2011).

Subsequently, RAN translation was also observed to occur in CGG repeats present in the 5' UTR of FMR1 in fragile X-associated tremor ataxia syndrome (FXTAS), producing polyglycine and polyalanine containing proteins (Todd et al., 2013). A product of RAN translation from FMR1 containing both polyglycine and a C terminal sequence translated from the DNA adjacent to the repeats was additionally found to be toxic in a *Drosophila* model, demonstrating the relevance of these RAN translation products to repeat pathogenesis (Todd et al., 2013).

The mechanism for RAN translation has not been characterised, however several potential processes have been proposed. Translation initiation may be assisted by the propensity of repetitive DNA to form hairpin loops, which are thought to promote translation initiation in some situations by slowing the 40S subunit of the ribosome, allowing tRNA anticodons to bind (Kozak, 1990, 1989). In support of this hypothesis, hairpin-forming CAG repeats were

found to be capable of initiating RAN translation, while non-hairpin-forming CAA repeats could not (Zu et al., 2011). A second potential mechanism for the initiation of RAN translation is the presence of near ATG codons, differing from ATG by only one base, which have been observed to initiate translation under certain conditions (Peabody, 1989), and were found to be required for RAN translation of repeats to occur within an *in vitro* system (Zu et al., 2011)

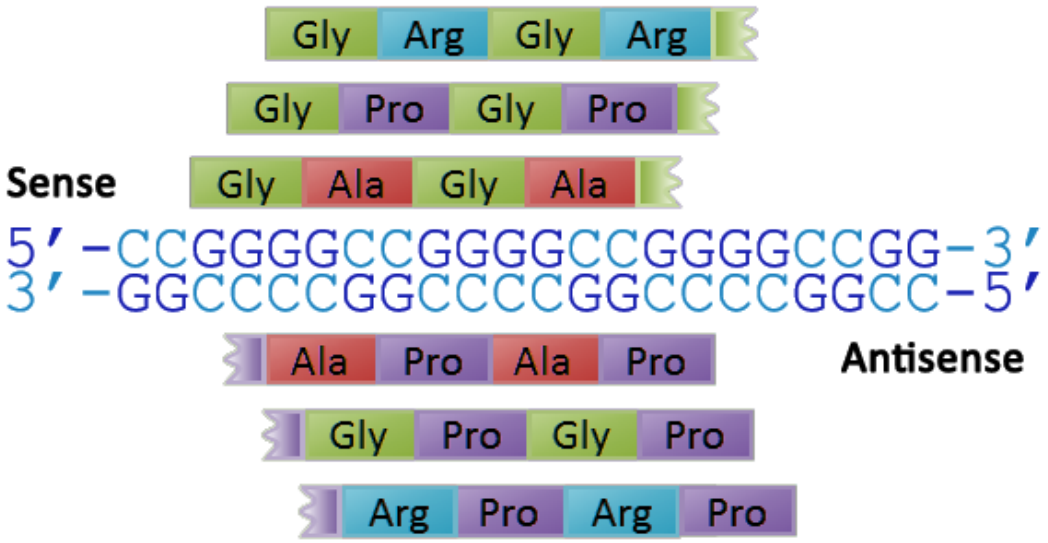


Figure 1.6 Dipeptide repeat protein RAN translation products from GGGGCC repeats

The dipeptide repeat protein products translated from the GGGGCC repeat expansion are shown for each reading frame in the sense and antisense orientations. gly: glycine, arg: arginine, pro: proline, ala: alanine.

In 2013, several groups demonstrated that RAN translation also occurs in *C9orf72* ALS/FTD. The GGGGCC repeat expansion is translated in all reading frames from both the sense (Ash et al., 2013; Mori et al., 2013c) and antisense transcripts (Gendron et al., 2013; Mori et al., 2013a; Zu et al., 2013). This yields 5 different dipeptide repeat (DPR) proteins: glycine-alanine (GA) and glycine-arginine (GR) in the sense orientation, alanine-proline (AP) and proline-arginine (PR) in the antisense orientation, and glycine-proline (GP) from one sense frame and one antisense frame (Figure 1.6). These proteins were found to be a component of p62 positive, TDP-43 negative inclusions that had previously been observed to be a

defining pathological hallmark of *C9orf72* FTD/ALS (Al-Sarraj et al., 2011; Mackenzie et al., 2013; Mann et al., 2013). DPR proteins form characteristic "star-like" cytoplasmic inclusions and small round intranuclear inclusions (Figure 1.7) in neurons from many different brain regions in *C9orf72* FTD/ALS patients, and are additionally observed in dystrophic neurites (Ash et al., 2013; Mackenzie et al., 2013; Schludi et al., 2015). DPR pathology is present most abundantly throughout the neocortex, hippocampus, cerebellum, and thalamus (Schludi et al., 2015). Less frequently, DPR proteins are also observed in other subcortical nuclei, some areas of the brain stem and in the spinal cord (Schludi et al., 2015). While DPR aggregates were originally purported to be present only in neurons (Mackenzie et al., 2013), recent studies have shown that this pathology can also be observed in the glial ependymal and subependymal cells that line the spinal cord central canal (Schludi et al., 2015). Aggregates of DPR products from translation of the sense strand are observed to be significantly more common than those from the antisense strand, with poly(GA) the most abundant protein, followed by less frequent inclusions of poly(GP) and poly(GR), while poly(AP) inclusions and poly(PR) inclusions are significantly less common (Mackenzie et al., 2015). The abundance of poly(GA) aggregates across all brain regions was observed to negatively correlate with age of disease onset in *C9orf72* FTD/ALS patients, (Davidson et al., 2014; Mackenzie et al., 2015), and the accumulation of DPR protein aggregates has been observed to predate TDP-43 pathology (Baborie et al., 2014; Gijselinck et al., 2012; Proudfoot et al., 2014), consistent with a role for DPR proteins in disease pathogenesis. However, there is no clear relationship between the frequency of DPR protein aggregates within any given brain region and the severity of neurodegeneration or TDP-43 pathology in that region (Davidson et al., 2014; Mackenzie et al., 2015, 2013; Mann et al., 2013), arguing against a direct relationship between DPR aggregate formation and cell toxicity. While several immunohistochemical studies have observed no difference in DPR aggregate distribution between *C9orf72* FTD, ALS or FTD/ALS patients (Davidson et al., 2014;

Mackenzie et al., 2015; Schludi et al., 2015), one study using immunoassays found significantly less poly(GP) in the cerebellum of *C9orf72* ALS patients compared to *C9orf72* patients with FTD or FTD/ALS (Gendron et al., 2015), and a second study also found a higher abundance of poly(GA) inclusions in the cerebellum granule cell layer of *C9orf72* FTD cases than in *C9orf72* ALS or FTD/ALS cases (Schludi et al., 2015), suggesting that differential DPR protein accumulation could potentially influence disease subtype.

The toxicity of several of the DPR proteins has also been demonstrated in a number of different cell and animal models. In several immortalised cell line models, toxicity mediated by the arginine-containing DPR proteins poly(GR) and poly(PR) has been observed (Kwon et al., 2014; Tao et al., 2015; Yamakawa et al., 2014). A reduction in cell line viability was also reported as a result of poly(GA) expression (Zhang et al., 2014). However, other studies did not observe toxicity as a result of any of these species (May et al., 2014). DPR protein toxicity has also been observed in rat primary neuronal cultures, with a reduction in cell viability caused by poly(GA) expression found by some groups (May et al., 2014; Zhang et

Figure 1.7 Dipeptide repeat protein inclusions in *C9orf72* FTD patient frontal cortex

Poly(glycine-alanine) aggregates in immunostained post-mortem *C9orf72* FTD/ALS patient frontal brain. Blue stain represents hematoxylin nuclear stain, brown stain represents 3,3' –diaminobenzidine stain against an anti-poly(glycine-alanine) primary antibody. A) Inclusions in frontal cortex. Black scale bar = 20 µm in part A. B) Star-like cytoplasmic inclusion in spinal cord Grey scale bar = 20 µm in part B. C) Cytoplasmic inclusion in the frontal cortex D) Dot-like nuclear inclusion in the frontal cortex E) Diffuse cytoplasmic staining usually defined as a "pre-inclusion" in the frontal cortex F) Dystrophic neurite in the frontal cortex. White scale bar = 20 µm in parts C-F. Adapted from (Mackenzie et al., 2015)

al., 2014) and toxicity mediated by arginine-containing DPR protein expression reported by others (Wen et al., 2014). Additionally, poly(GR) and poly(PR) were both found to cause toxicity in neurons differentiated from human iPSCs (Wen et al., 2014).

When expressed in *Drosophila*, poly(GR) and poly(PR) proteins were sufficient to cause severe toxicity (Mizielinska et al., 2014; Wen et al., 2014; Yang et al., 2015). Poly(GR) expression was additionally found to suppress Notch signalling in flies (Yang et al., 2015), suggesting that these peptides could disrupt key cell signalling pathways in disease. A modest reduction in viability was also observed in poly(GA) expressing *Drosophila* (Mizielinska et al., 2014), however this was found to be considerably less severe than the toxicity caused by poly(GR) and poly(PR).

1.6.4 Proposed cellular functions affected in *C9orf72* FTD/ALS

A number of cellular functions are reported to be compromised as a result of gain-of-function mechanisms initiated by the *C9orf72* repeat expansion. Firstly, several groups have highlighted the effect of arginine-containing DPR proteins on nucleoli in cell model systems, following initial observations that poly(GR) and poly(PR) peptides localise to the nucleolus (Kwon et al., 2014; May et al., 2014; Schludi et al., 2015; Tao et al., 2015; Wen et al., 2014; Yamakawa et al., 2014; Zu et al., 2013). The presence of poly(GR) and poly(GR) in cell culture models was also found to induce nucleolar enlargement, translocation of nucleolar markers and disrupted processing of ribosomal subunits (Kwon et al., 2014; Tao et al., 2015; Wen et al., 2014). Other groups have additionally implicated RNA gain-of-function mechanisms in nucleolar dysfunction (Haeusler et al., 2014). However thus far studies have failed to observe the same changes in nucleolar morphology in *C9orf72* FTD/ALS post mortem brain (Schludi et al., 2015). The role of the nucleolus in *C9orf72* repeat pathology will be discussed in greater detail in chapter 5.

Secondly, a number of reports have highlighted transcriptome changes in *C9orf72* ALS patients and patient derived cells (Cooper-Knock et al., 2015a; Donnelly et al., 2013; Lagier-Tourenne et al., 2010; Prudencio et al., 2015; Sareen et al., 2013). Alterations in alternative splicing and alternative polyadenylation have been observed in *C9orf72* ALS patient tissue and cells, but the exact transcripts found to be dysregulated has been inconsistent between studies (Cooper-Knock et al., 2015a; Donnelly et al., 2013; Lagier-Tourenne et al., 2010; Prudencio et al., 2015; Sareen et al., 2013), which may in part be due to a reduction of splicing consistency in these samples (Cooper-Knock et al., 2015a). Additionally, transcripts in *C9orf72* ALS patient tissue and cells were observed to significantly deviate not only from that of control patients, but also from non-*C9orf72* ALS patients, (Cooper-Knock et al., 2015a; Donnelly et al., 2013; Prudencio et al., 2015), indicating that the disruption of RNA metabolism seen in *C9orf72* ALS patients may differ from other ALS patients in many respects. However, it has not yet been determined whether the observed dysregulation of RNA metabolism is caused by RNA or DPR protein gain-of-function mechanisms.

Finally, three recent papers observed defects in nucleocytoplasmic transport in transgenic GGGGCC repeat expressing *Drosophila* models, in a PR_{50} expressing yeast line, and in *C9orf72* FTD patient derived iPSCs, with both RNA gain-of-function and DPR protein gain-of-function mechanisms implicated by different groups (Freibaum et al., 2015; Jovičić et al., 2015; Zhang et al., 2015). The potential contribution of a defect in nucleocytoplasmic transport to *C9orf72* FTD/ALS and the relevance of these new developments to this thesis will be examined in depth in section 6.3 of the final discussion chapter.

Further studies are therefore needed to confirm the involvement of all of the above processes in *C9orf72* FTD/ALS, and how they may be impacted by RNA or DPR protein gain-of-function mechanisms.

1.7 Thesis aims

The *C9orf72* repeat expansion results in highly diverse clinical outcomes, with an array of different types of neuropathology observed in patients. It is presently unclear which pathological features play an important role in disease progression, which determine clinical subtype, and which may be epiphenomena playing a neutral or even protective role. A better understanding of which molecular species mediate toxicity in *C9orf72* FTD/ALS and by what mechanisms would assist in targeting future treatments for this disease.

This project therefore sought to dissect the different *C9orf72* FTD/ALS disease mechanisms at work, to study their effects in isolation. As the current evidence outlined above supports a gain-of-function mechanism as the primary driver of *C9orf72* FTD/ALS pathology, I will principally be focussing on dissecting the two major gain-of function processes: repeat RNA gain-of-function and DPR protein gain-of function.

My principal aims are:

- To generate a genetic toolkit that can be used to dissect RNA and DPR protein gain-of-function mechanisms.
- To use this toolkit to study gain-of-function disease mechanisms in cell model system.
- To study the relevance of these gain-of-function mechanisms in *C9orf72* FTD patient tissue.

In order to achieve these aims, I carried out the following studies:

- Generation of a tool kit of DNA repeat constructs of differing lengths that isolate RNA gain-of-function from DPR-protein gain-of-function (Chapters 3 and 4).

- Quantification of the formation of sense and antisense RNA foci from different length repeat constructs in a cell model system (Chapter 3).
- Investigation into the different effects of RNA gain-of-function and DPR-protein gain-of-function mechanisms in a transgenic *Drosophila* model generated using my repeat constructs (Chapters 3 and 4).
- Examination of the subcellular location of DPR proteins generated from my repeat constructs in a cell model system (Chapter 4)
- Development and testing of an automated image analysis protocol for the quantification of DPR protein inclusions in *C9orf72* FTD post mortem brain (Chapter 4).
- Investigation into a potential pathological mechanism at work in *C9orf72* FTD patients by examining the effect of poly(GR) inclusion burden on nucleolar volume in *C9orf72* FTD post mortem brain (Chapter 5).

Chapter 2: Materials and Methods

2.1 Cloning Protocols

2.1.1 Annealing and ligation of oligonucleotides

Poly-acrylamide gel electrophoresis (PAGE) purified oligonucleotides were ordered from Sigma Aldrich (Table 2.1). Forward and reverse oligonucleotides were diluted to 100 µM in annealing buffer (1 mL annealing buffer stock solution: 100 µL 1M Tris pH8 (Sigma), 500 µL 1M NaCl (VWR), 20 µL 0.5 M EDTA (Sigma), 280 µL ddH_2O, stored at -20 °C). The reaction was then heated to 95 °C for 15 minutes, after which the temperature was decreased by 0.1 °C per minute until the temperature reached 4 °C. To ligate the annealed oligonucleotides into a vector, 7.5 µL of oligonucleotides diluted to 7.5 nM and 0.5 µL of 20 ng/µL linearised vector with compatible overhangs were incubated with 1 µL 10X T4 ligase buffer and 1 µL T4 DNA ligase (40 U, New England Biolabs) at 16 °C for 2 hours.

2.1.2 Preparation of bacterial culture solutions

2.1.2.1 Luria-Broth (LB; Miller's Modification)

To make 500 mL of Luria-Broth, 12.5g LB powder (Sigma) was added to 500 mL ddH_2O, and was autoclaved at 121 °C for 15 minutes to sterilise. The solution was then left to cool to room temperature before addition of either 1) 100 µg/mL ampicillin (for pBlueScript II SK+, pcDNA3.1+ and pUASTattB vectors) or 2) 50 µg/mL kanamycin (for pAcGFP1 vector). The solution was stored sealed at room temperature until use.

2.1.2.2 LB agar

To make 500 mL of LB agar, 15.25 g LB Agar powder (Sigma) was added to 500 mL ddH_2O and was autoclaved at 121 °C for 15 minutes to sterilise. The solution was then left to cool until approximately 40 °C before addition of either 1) 100 µg/mL ampicillin (for pBlueScript

Insert	Destination vector	Sequence (5' to 3')
$(GGGGCC)_3$ forward	pcDNA3.1+	GATCCGGTACCGCTCTTCAGGCCGGGGCCGGGGCCGGGGCCTGC
$(GGGGCC)_3$ reverse	pcDNA3.1+	GCCATGGCGAGAAGTCCGGCCCCGGCCCCGGCCCCGGACGCCGG
$(GGGGCC)_4$ + int A forward	pBlueScript II SK+	[Phos]GGCCAAATTAGGGGCCGGGGCCGGGGCCGG
$(GGGGCC)_4$ + int A reverse	pBlueScript II SK+	[Phos]GCCCCGGCCCCGGCCCCGGCCCCTAATTTG
$(GGGGCC)_4$ + int B forward	pBlueScript II SK+	[Phos]GGCCGTTAAAGGGGCCGGGGCCGGGGCCGG
$(GGGGCC)_4$ + int B reverse	pBlueScript II SK+	[Phos]GCCCCGGCCCCGGCCCCGGCCCCTTTAACG
$(GGGGCC)_4$ + int C forward	pBlueScript II SK+	[Phos]GGCCAATTAAGGGGCCGGGGCCGGGGCCGG
$(GGGGCC)_4$ + int C reverse	pBlueScript II SK+	[Phos]GCCCCGGCCCCGGCCCCGGCCCCTTAATTG
$(Gly\text{-}Ala)_6$ forward	pAcGFP1-C1	AATTCTGGTGCAGGAGCTGGTGCAGGAGCTGGTGCAGGAGCTG
$(Gly\text{-}Ala)_6$ reverse	pAcGFP1-C1	GGATCCAGCTCCTGCACCAGCTCCTGCACCAGCTCCTGCACCAG
$(Gly\text{-}Pro)_6$ forward	pAcGFP1-C1	AATTCTGGTCCAGGACCTGGTCCAGGACCTGGTCCAGGACCTG
$(Gly\text{-}Pro)_6$ reverse	pAcGFP1-C1	GGATCCAGGTCCTGGACCAGGTCCTGGACCAGGTCCTGGACCAG
$(Gly\text{-}Arg)_6$ forward	pAcGFP1-C1	AATTCTGGTAGAGGAAGAGGTAGAGGAAGAGGTAGAGGAAGAG
$(Gly\text{-}Arg)_6$ reverse	pAcGFP1-C1	GGATCCTCTTCCTCTACCTCTTCCTCTACCTCTTCCTCTACCAG
$(Pro\text{-}Arg)_6$ forward	pAcGFP1-C1	AATTCTCCTAGACCAAGACCTAGACCAAGACCTAGACCAAGAG
$(Pro\text{-}Arg)_6$ reverse	pAcGFP1-C1	GGATCCTCTTGGTCTAGGTCTTGGTCTAGGTCTTGGTCTAGGAG
$(Ala\text{-}Pro)_6$ forward	pAcGFP1-C1	AATTCTGCTCCAGCACCTGCTCCAGCACCTGCTCCAGCACCTG
$(Ala\text{-}Pro)_6$ reverse	pAcGFP1-C1	GGATCCAGGTGCTGGAGCAGGTGCTGGAGCAGGTGCTGGAGCAG

Table 2.1 Oligonucleotides used for cloning GGGGCC and DPR protein repeats.

int = interruption, gly = glycine, pro = proline, arg = arginine, ala = alanine.

II SK+, pcDNA3.1+ and pUASTattB vectors) or 2) 50 µg/mL kanamyacin (for pAcGFP1 vector). LB agar solution was then poured into 10 cm petri dishes close to a flame, with 20 mL of solution added per dish. The agar was allowed to set before covering and storing at 4 °C.

2.1.3 Transformation of chemically competent *E. coli*

3 µL of ligation reaction was added to One Shot® Stbl3™ Chemically Competent *E. coli* (Life Technologies), a *recA* recombinase-deficient strain that promotes stability when cloning repetitive DNA. The cells were incubated on ice for 30 minutes before heat shock at 42 °C for 45 seconds, followed by incubation on ice for a further 2 minutes. 250 µL of pre-warmed SOC media was added to cells, which were then incubated at 30 °C shaking for 2 hours. Cells were incubated at 30 °C as opposed to the 37 °C incubation temperature recommended by manufacturers as the lower temperature was found to reduce the frequency of deletions in the repeat constructs by the bacteria. Cultures were then spread onto LB agar plates containing the appropriate antibiotic. Agar plates were incubated at 30 °C overnight.

2.1.4 Colony PCR for detection of GGGGCC repeats

Colonies of transformed *E. coli* on LB agar plates were each picked into 40 µL ddH_2O in a 96-well plate. 10 µL of these solutions was then added to 200 µL each LB broth with the appropriate antibiotic in a 96-well round bottom culture plate, which was then incubated shaking at 30 °C until completion of the PCR. The remaining colony solutions in ddH_2O were heated at 95 °C for 10 minutes. A repeat-primed PCR (Beck et al., 2013; Renton et al., 2011a) reaction master mix was made by multiplying the volumes listed in Table 2.2 by the number of colonies to be tested. 1 µL of solution from each well of the colony solutions in ddH_2O was then added to 10 µL of master mix per well. The forward and reverse primers used are detailed in Table 2.3, and the reaction was incubated as detailed in Table 2.4.

Reagent	Supplier	Volume per well (µL)
2x FastStart PCR Master Mix	Roche	5
DMSO	Sigma	0.5
5x Q solution	Qiagen	2
5 mM deazaGTP	New England Biolabs	0.4
25 mM MgCl2	Qiagen	0.4
100 µM forward primer	Sigma	0.05
100 µM reverse primer	Sigma	0.05
H_2O	-	0.6
DNA	-	1

Table 2.2 Reagents for colony PCR reaction

Primer	Sequence (5' to 3')
pBlueScript II SK+ forward (FAM labelled)	[6FAM]CCTCGAGGTCGACGGTA
pBlueScript II SK+ reverse	CGCGCAATTAACCCTCACTA

Table 2.3 Colony PCR primers

Step	Temperature	Time
1	95 °C	5 minutes
2	95 °C	30 seconds
3	50 °C	30 seconds
4	72 °C	1 minute
Repeat steps 2-4, 35 times		
5	72 °C	7 minutes
6	4 °C	Forever

Table 2.4 Cycling conditions for colony PCR reaction

1 µL of each PCR reaction to be analysed was mixed with 10 µL Hi-Di™ formamide (Life Technologies) and 0.5 µL GeneScan™ 500 LIZ® Size Standard (Life Technologies) in a 96-well plate, and fragment analysis was carried out using the 3730xl DNA Analyser (AME Bioscience), and Peak Scanner™ Software (Applied Biosystems).

The 200 µL LB starter cultures from samples that contained the correct length inserts (as indicated by fragment analysis) were added to 5 mL each LB broth with appropriate antibiotics. Cultures were incubated overnight shaking at 30 °C for plasmid DNA preparation by miniprep the following day.

2.1.5 Preparation of plasmid DNA

To seed bacterial cultures for plasmid minipreps, transformed *Stbl3* colonies on LB agar plates were picked into 5 ml LB broth with the appropriate antibiotic, followed by incubation shaking at 30 °C overnight. Due to the high repeat deletion rate in these constructs, 24 or 48 colonies were typically screened. DNA was extracted using the QIAprep Spin Miniprep kit (QIAGEN) as per the manufacturer's instructions. To elute, 50 µL of Buffer EB was added to the column membrane and left to stand for one minute before the final centrifugation step.

To prepare higher concentration and higher purity plasmid stocks for applications such as cell transfection, plasmid maxipreps were carried out. First, minipreps were carried out as detailed above, however 0.5 ml of each LB culture was retained prior to DNA extraction, and was added to 4 ml LB broth containing the appropriate antibiotic, and incubated shaking at 30 °C for a further 6 hours in order to seed maxiprep starter cultures. DNA minipreps were digested (Section 2.1.7) to release the repeat insert and screen for correct size by agarose gel electrophoresis (Section 2.1.8). 500 µL of starter culture corresponding to a successful prep was then added to 500 ml of LB broth with the appropriate antibiotic, and was incubated shaking overnight at 30 °C. The culture was processed using the

EndoFree Maxi Prep Kit (QIAGEN) as per the manufacturer's instructions. The final DNA pellet was left in 500 µL Buffer TE (QIAGEN) overnight to solubilise.

2.1.6 DNA UV spectroscopy

Concentration of DNA samples was determined by loading 1 µL of solution onto a NanoDrop ND-1000 spectrophotometer (Thermo Scientific) and measuring absorbance at 260 nm. 1 µL of the elution solution used (either ddH_2O, Buffer EB or Buffer TE depending on the experiment), was used as a blank reference sample to calibrate the spectrophotometer. Purity of the DNA sample was checked using the 260 nm to 280 nm ratio.

Purpose	Digest	Units of enzyme used per 50 µL reaction	Buffer	Temperature
Excise repeat insert for RDL	BspQI/EcoO10I double digest	BspQI: 10 U EcoO10I: 20 U	CutSmart Buffer	37 °C for 1 hour 50 °C for 1 hour
Linearise repeat-containing vector for RDL	BspQI single digest + Calf Intestinal Phosphotase (CIP) treatment	BspQI: 10 U CIP: 10 U	CutSmart Buffer	50 °C for 1 hour, CIP added, 37 °C for 1 hour
Excise repeats for subcloning into pcDNA3.1 vector	BamHI-HF/NotI-HF Double Digest	BamHI-HF: 20 U NotI-HF: 20 U	CutSmart Buffer	37 °C for 2 hours
Excise repeats for subcloning into pUAST attb vector	EcoRI-HF/NotI-HF	NotI-HF: 20 U EcoRI-HF: 20 U	CutSmart Buffer	37 °C for 2 hours
Linearise pAcGFP1-C1 vector for ligation with $(DPR)_6$ oligos	EcoRI-HF/BamHI-HF	EcoRI-HF: 20 U BamHI-HF: 20 U	CutSmart Buffer	37 °C for 2 hours

Table 2.5 Restriction enzyme digests.

RDL = Recursive directional ligation.

2.1.7 Restriction enzyme digestion

Reactions were set up as in Table 2.5 and were incubated in a DNA Engine Tetrad®2 Thermal Cycler (BioRad) at the required temperature and duration. Single restriction enzyme digestion reaction products were treated with Calf Intestinal Phosphotase (CIP) to dephosphorylate the free DNA ends and prevent relegation. All restriction enzymes, phosphotases and digestion buffers were purchased from New England Biolabs.

2.1.8 Agarose gel electrophoresis

Between 0.8 % and 4 % UltraPure™ Agarose powder (Life technologies) was added to 200 mL 1 x TBE, depending on the predicted size of the digest product of interest. Very small digestion products smaller than 200 bp were resolved on a 4 % gel, digestion products between approximately 200 bp and 1 kb were resolved on a 2 % gel, and larger digestion products above 1 kb in length were resolved on a 1 % or 0.8 % gel. The gel mixture was heated in a microwave for 2 minutes and left to cool to approximately 50 °C before the addition of 20 µL Ethidium Bromide solution (500 µg/ml in H_2O; Sigma) per 200 mL gel, inside a chemical fume hood. The gel solution was then poured into a holder with a lane comb and left to set. Samples were mixed with 5x DNA loading buffer Blue (Bioline), before loading into the gel. A molecular weight marker was also loaded to allow approximation of digestion band size, with Hyperladder I (Bioline) loaded for bands of interest up to 1 kb in length, and Hyperladder IV (Bioline) loaded for bands of interest larger than 1 kb. Electrophoresis was carried out in 1 x TBE buffer at 120 V for 30-90 minutes. The gel was imaged using a Gel Doc XR system with Quantity One software (BioRad).

2.1.9 DNA gel extraction

The bands of interest were visualised on a UV transilluminator (Peqlab biotechnology GmbH) and were excised from the agarose gel with a scalpel, with care taken to limit the exposure of the samples to the UV radiation. DNA was extracted using the QIAquick Gel

extraction kit (QIAGEN) as per manufacturer's instructions. To elute, 30 µL of ddH_2O was added to the column membrane and left to stand for one minute before the final centrifugation step.

2.1.10 DNA Ligation

A volume of insert and vector DNA to be ligated totalling 8 µL was incubated with 1 µL 10X T4 ligase buffer and 1 µL T4 DNA ligase (40 U, New England Biolabs) at 16 °C overnight. As the ligation of multiple inserts during recursive directional ligation cloning (see section 3.2.1) was advantageous, a high insert to vector ratio was used, between 10:1 and 100:1 molar ratio depending on the concentration of DNA available. For subcloning, lower ratios between 3:1 and 10:1 were more typically used.

2.1.11 Sequencing

Constructs to be sequenced were either processed by Source Bioscience using dGTP chemistry to aid read-through of GC-rich repeats, or were sequenced as follows: The reagents listed in Table 2.6 were added for each sequencing reaction in a 96-well plate. 1 µL of a 5 pmol/µL stock of the required primer was added. For sequencing pBlueScript II SK+ constructs either T7 forward primer or M13 reverse primer was added. For pcDNA3.1+, pUASTattB or pAcGFP1-C1 primers, see Table 2.7. A PCR amplification reaction was then carried out as detailed in Table 2.8.

Reagent	Supplier	Volume per well (µL)
Big dye	Life Technologies	1
Better buffer	Life Technologies	5
Betaine	Sigma	1
Primer	Sigma	1
ddH_2O		6.5
DNA		1.5

Table 2.6 Reagents for sequencing reaction

Primer	Sequence (5' to 3')
pcDNA3.1+ forward	ATCAACGGGACTTTCCAAAATGTCG
pcDNA3.1+ reverse	TGACACCTACTCAGACAATGCGATG
pUASTattB forward	GCGAAAGCTAAGCAAATAAACAAGC
pUASTattB reverse	CCCATTCATCAGTTCCATAGGTTGG
pAcGFP1-C1 forward	ACTACCTGTCCACCCAGAGC
pAcGFP1-C1 reverse	TGAGTTTGGACAAACCACAAC
T7 forward	TAATACGACTCACTATAGGG
M13 reverse	CAGGAAACAGCTATGAC

Table 2.7 Sequencing primers

Step	Temperature	Time
1	96 °C	30 seconds
2	50 °C	15 seconds
3	60 °C	3 minutes
Repeat steps 1-3, 30 times		
4	15 °C	5 mins

Table 2.8 Cycling conditions for sequencing PCR reaction

DNA was precipitated by incubation at room temperature with 1.5 µL 125 mM EDTA (Sigma), 1.5 µL 3M sodium acetate (Sigma) and 37.5 µL 100 % ethanol per well, followed by centrifugation at 2500 rcf for 30 minutes. Supernatant was removed, and 70 µL of 70 % ethanol was added per well. Samples were centrifuged for 15 minutes at 2000 rcf, supernatant was removed, and samples were left to dry for 10 minutes. 10 µL of Hi-Di formamide (Life Technologies) was added to each well, and the sample was sequenced using the 3730xl DNA Analyser (AME Bioscience, Life Technologies).

2.1.12 Plasmids

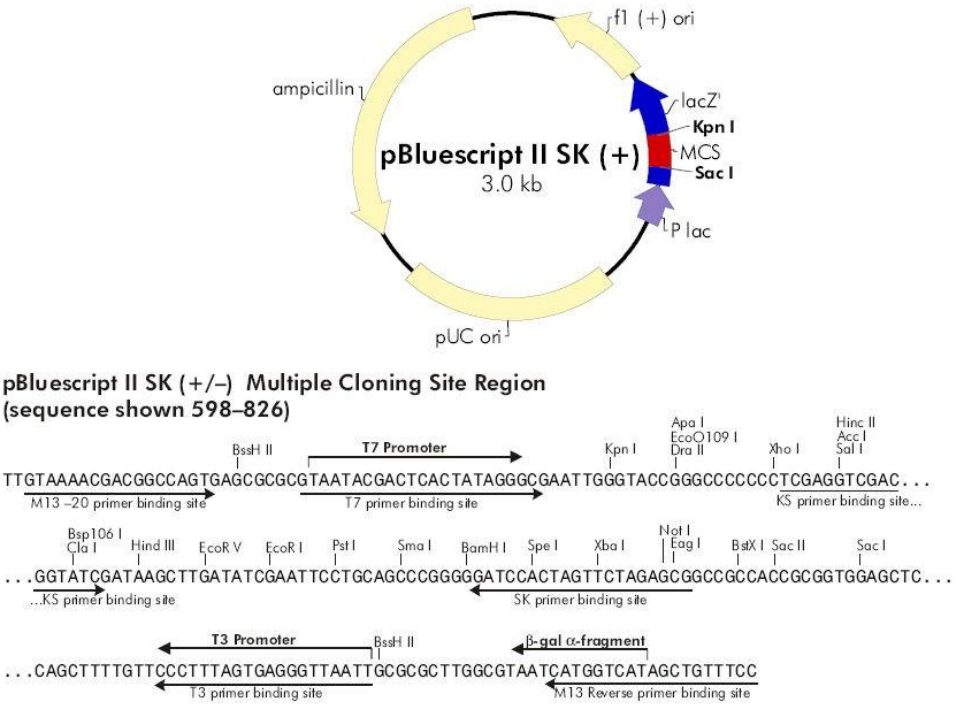


Figure 2.1 pBlueScript II SK+ cloning vector (Agilent #212205)

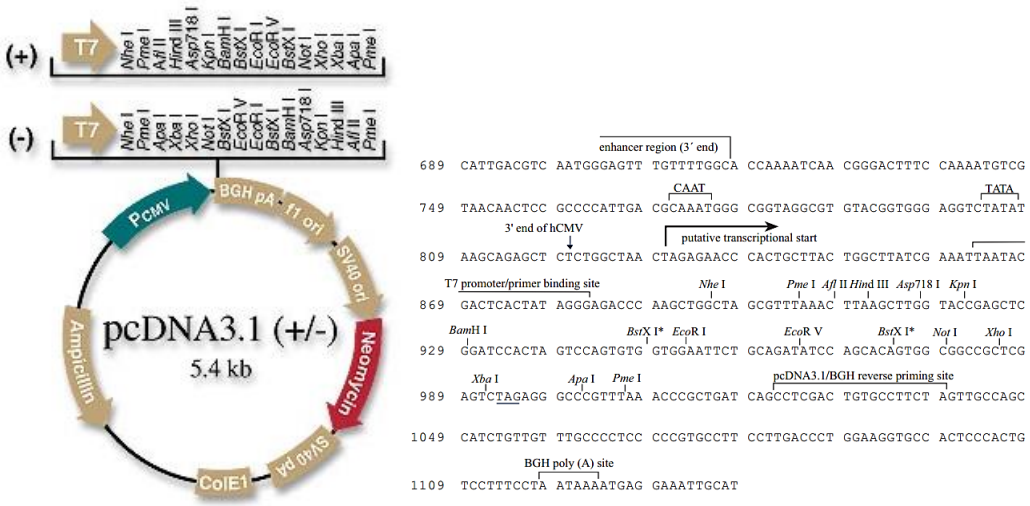


Figure 2.2 PCDNA3.1+ mammalian expression vector (Life Technologies V795-20)

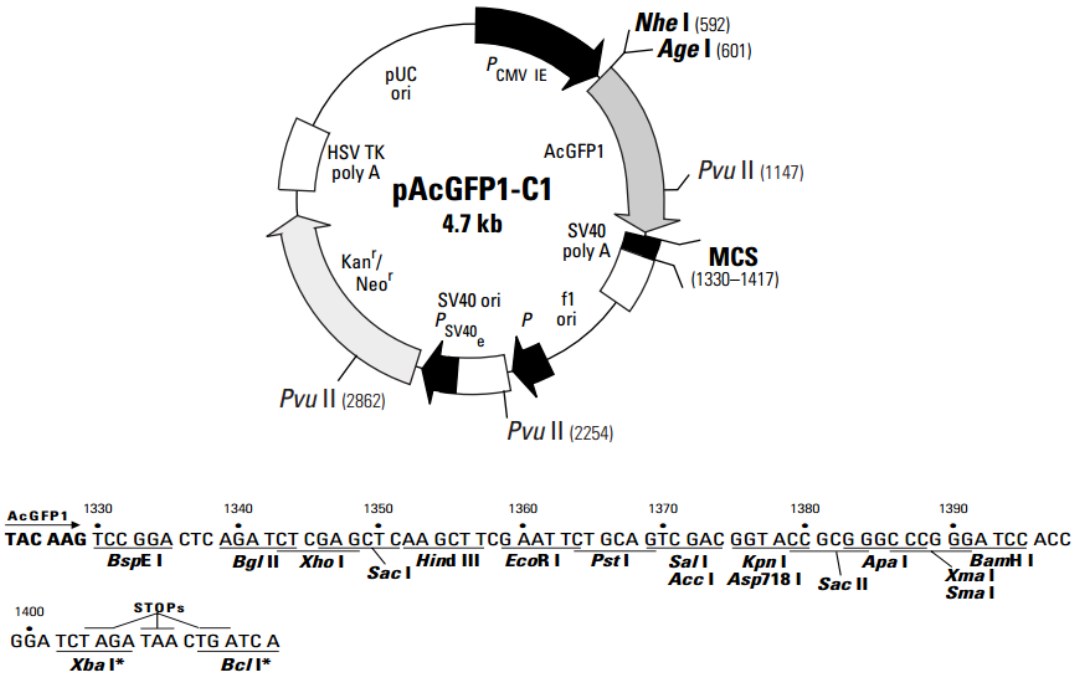


Figure 2.3 pAcGFP1-C1 mammalian expression vector (Clontech 632470)

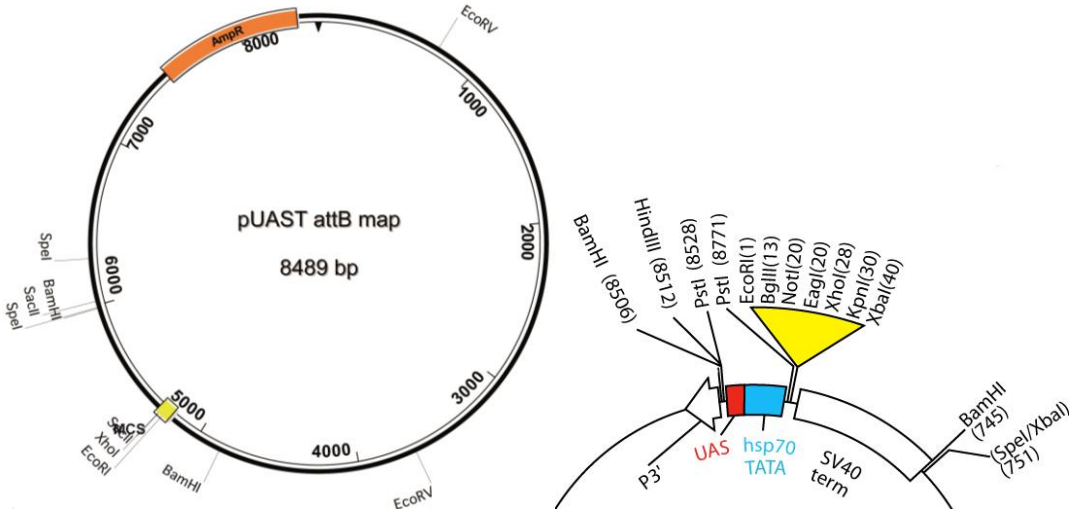


Figure 2.4 pUASTattB Drosophila expression vector

2.2 Cell culture

2.2.1 Culture and passage

SH-SY5Y cells and HEK 293T cells were grown in complete DMEM (DMEM cat #11960-085 with 10 % FCS, 5 mM L-Glutamine, 1 mM Pyruvate; Life Technologies), in 10 cm dishes incubated at 37 °C/5 % CO_2. To passage, cells were washed with 10 mL PBS then incubated in 1 mL trypsin with EDTA (Life Technologies) for 2 minutes at 37 °C to detach cells from the dish. The reaction was stopped with the addition of 9 mL complete DMEM. Cells were dislodged from the dish, transferred into a 15 mL falcon, and were pelleted at 1000 rpm for 5 minutes. The supernatant was discarded and cells were resuspended in 10 mL complete DMEM. Different volumes of this cell solution were then added to several new 10 cm dishes and made up to a final volume of 10 mL with complete DMEM to seed several new cell cultures at varying concentrations for continued growth.

2.2.2 Cryopreservation of cells

For long-term storage, early passage cells from a near confluent dish were typsinised, resuspended in complete DMEM, and spun down as detailed in the previous section. After removal of the supernatant, the cell pellet was resuspended in 1 mL freezing media, which comprised of 90 % complete DMEM culture media with 10 % dimethyl sulphoxide (DMSO; Sigma), and transferred into a cryovial. Cryovials were then placed into a Mr. Frosty™ Freezing Container (Thermo) filled with isopropanol for gradual cooling, and were transferred into a -70 °C freezer overnight. Once thoroughly frozen, the cryovials were moved into a liquid nitrogen tank for long term storage.

2.2.3 Revival of cryopreserved cell stocks

To defrost cells from a liquid nitrogen stock, individual cryovials were rapidly thawed in a 37 °C waterbath. 1 mL of prewarmed complete DMEM was added in a dropwise manner to

the cryovial while constantly agitating, before transferring into a 15 mL falcon. A further 8 mL of complete DMEM was added to the falcon tube while constantly agitating, before pelleting cells at 1000 rpm for 5 minutes. The cell pellet was then resuspended in 10 mL complete DMEM, transferred into a 10 cm dish and incubated at 37 °C/5 % CO_2. Defrosted cells were grown and passaged as detailed above for at least one week before use in any experiments.

2.2.4 Transient transfection

SH-SY5Y or HEK 293T cells were plated on 13 mm diameter round glass coverslips in a 24 well dish at a density of 1.5 x 10^{-4} cells per well, in 500 µL complete DMEM and were incubated at 37 °C/5 % CO_2 overnight. For each well, 0.4 ng of DNA to be transfected was mixed with 0.04 ng of co-transfection marker (ZsGreen or DsRed; Clontech), and 30 µL of OptiMEM (Life Technologies). 1 µL Lipofectamine 2000 (Life Technologies) diluted in 30 µL OptiMEM was then added, and the solution was incubated at room temperature for 20 minutes. In this time, media on the cells was replaced with fresh complete DMEM. The transfection solutions were then added dropwise to the wells, and were left to incubate at 37 °C, 5 % CO2 for 3 hours. The cell culture media then was replaced with fresh complete DMEM, and cells were incubated for a further 24 hours at 37 °C, 5 % CO_2.

2.2.5 Fluorescence *in situ* hybridisation

Transfected cells were washed in PBS, fixed in 4 % paraformaldehyde (PFA, Polysciences) in PBS for 10 mins, washed briefly in 2 x SCC (Sigma), and then incubated in pre-hybridisation solution (40 % formamide; Amresco/2 x SSC) for 30 minutes at room temperature. Cells were then incubated for 2 hours at 37 °C in hybridisation solution with either a Cy3-labelled $(GGCCCC)_4$ probe (Integrated DNA Technologies) for sense RNA foci detection or an Alexa 488-labelled $(GGGGCC)_4$ probe (Integrated DNA Technologies) for antisense RNA foci detection. Hybridisation solution consisted of: 0.2 % BSA (Sigma), 2 x SSC, 1 mg/mL salmon

sperm DNA (Sigma), 1 mg/mL tRNA (Sigma), 40 % formamide, 10 % dextran sulphate (Sigma), 2 mM vanadyl ribonucleoside (New England Biolabs), with either 20 ng/µL Cy3-labelled $(GGCCCC)_4$ probe or 25 ng/µL Alexa 488-labelled $(GGGGCC)_4$ probe. Samples were then shaken at 37 °C for 3 x 20 minutes in wash solution (40 % formamide/ 2 x SCC), followed by 3 x 10 minute washes in 2x SCC at room temperature. Cells were mounted using ProLong Gold antifade reagent with DAPI to visualise cell nuclei (Life Technologies).

2.2.6 Immunofluorescence

Transfected cells were washed in PBS before fixation in 4 % PFA for 10 minutes at room temperature. At this point the cells could be stored in PBS for up to a month if desired. Cells were permeabilised with 0.1 % Triton X-100 in PBS for 10 minutes, before incubation in 1 % BSA in PBS for 30 minutes to block non-specific interactions. Coverslips were then inverted onto a drop of the relevant primary antibodies diluted in PBS (see Table 2.9 below for details), and were left to incubate at room temperature for 1 hour. Cells were then transferred back to the 24-well dish and washed for 3 x 5 minutes in PBS, before the coverslips were inverted onto a drop of the relevant secondary antibodies diluted in PBS (see Table 2.10 below for details), and incubated for 1 hour. Finally the cells were washed as before for a further 3 x 5 minutes in PBS before mounting on a glass microscope slide using ProLong Gold Antifade Mountant with DAPI.

2.3 Antibodies

2.3.1 Generation of antibodies to dipeptide repeat proteins

Custom antibodies to each dipeptide repeat protein were generated by BioGenes GmbH. Rabbits were immunised with fusion proteins containing 7 dipeptide repeats (GR, GP, GA, AP or PR) conjugated to limulus polyphemus hemocyanin (LPH).

Target	Host	Type	Source	Product code	Application	Dilution
Nucleophosmin	Mouse	IgG_1	Abcam	ab10530	Human IF	1:1000
Nucleolin	Mouse	IgG_1	Santa Cruz	sc-8031	Human IF	1:50
NeuN	Rabbit	IgG	Merck Millipore	ABN78	Human IF	1:1000
GR	Rat	Hybridoma supernatant	Kind gift from Friedrich Grässer	5H9	Human IF	1:25
					Human IHC	1:50
GR	Rabbit	IgG	Proteintech	23978-1-AP	Cell IF	1:200
GA	Rabbit	IgG	Proteintech	24492-1-AP	Cell IF	1:200
GA	Rabbit	Antiserum	Biogenes*	N/A	Human IHC	1:200
GP	Rabbit	Antiserum	Biogenes*	N/A	Cell IF	1:200
					Human IHC	1:4000
PR	Rabbit	IgG	Kind gift from Stuart Pickering-Brown	N/A	Cell IF	1:50
PR	Rabbit	IgG	Biogenes*	N/A	Human IHC	1:200
AP	Rabbit	Antiserum	Biogenes*	N/A	Cell IF	1:200
					Human IHC	1:50

Table 2.9 Primary antibodies used
* See section 2.3.1 for manufacture details of antibodies generated by Biogenes. IF = Immunofluorescence IHC = Immunohistochemistry

Target	Host	Conjugate	Manufacturer	Product code	Application	Dilution
Mouse	Goat	Alexa 488	Life Technologies	A11029	Cell IF / Human IF	1:500
Rat		Alexa 546		A11081		
Rabbit		Alexa 546		A11035		
Rabbit		Alexa 633		A21071		
Chicken		Alexa 633		A21103		
Rabbit	Swine	Biotin	Dako	E035301-2	Human IHC	1:200

Table 2.10 Secondary antibodies used
IF = Immunofluorescence IHC = Immunohistochemistry

2.4 Human histology

2.4.1 Cases

Case number	PM delay (hrs:mins)	AAD	Gender	Clinical Diagnosis	Cause of Death	Path Diagnosis
1	171:00	69	M	Neurologically normal	Myocardial infarction	Normal/ path ageing
2	24:00	73	F	Neurologically normal	Pulmonary embolism	Path ageing
3	16:15	88	M	Neurologically normal	Cardiac arrest	Path ageing/ TDP-43
4	99:00	83	F	Neurologically normal	Bronchopneumonia/ renal failure	Normal/ path ageing/ incidental lewy bodies
5	49:10	80	F	Neurologically normal	Pancreatic carcinoma	Normal/ path ageing
6	120:00	86	F	Neurologically normal	N/a	Normal/ path ageing
7	29:40	93	F	Neurologically normal	N/a	Normal/ path ageing/ TDP-43
8	43:07	91	F	Neurologically normal	Cerebrovascular disease	Normal/ path ageing/ incidental lewy bodies

Table 2.11 Control patient post mortem case details

PM = post mortem, AAO = age at onset, AAD = age at death Path = Pathological

Case number	PM delay (hrs:mins)	AAO	AAD	Duration	Gender	Clinical Diagnosis	Path Diagnosis	Mutations
9	85:35	56	67	11	F	HD (EARLY ONSET)	FTLD-TDP, Type A	*C9orf72*
10	99:00	62	68	6	M	FTD	FTLD-TDP type A	*C9orf72*
11	51:52	66	71	5	M	FTD/MND	FTLD-TDPA	*C9orf72*
12	30:00	59	65	6	M	FTD	FTLD-TDPA	*C9orf72*
13	115:00	58	66.8	8.8	F	MND	FTLD-TDP type A	*GRN*/ *C9orf72*
14	94:05	64	66	2	F	FTD	FTLD-TDP, type B	*C9orf72*
15	63:05	57	62	5	F	FTD	FTLD-TDP	*C9orf72*
16	32:20	54	60	6	M	FTD	FTLD-TDP, type A	*C9orf72*
17	77:20	53	63	10	M	FTD	FTLD-TDP, type A	*C9orf72*
18	85:50	66	74	8	F	FTD	FTLD-TDP type A	*C9orf72*
19	25:53	43	45	2	M	FTD	FTLD-TDP, Type A	*C9orf72* homozygous

Table 2.12 *C9orf72* patient post mortem case details.

PM = post mortem, AAO = age at onset, AAD = age at death Path = Pathological

Paraffin embedding (Section 2.4.2), sectioning (Section 2.4.2) and immunohistochemical staining for light microscopy (Section 2.4.3), of human brain tissue was kindly carried out by Dr. Tammaryn Lashley and colleagues at the Queen Square Brain Bank. Presence of the *C9orf72* repeat expansion in FTD cases was confirmed by Southern blot or repeat primed PCR.

2.4.2 Tissue Processing

Formalin fixed brain tissue was embedded in paraffin wax by dehydration in a series of ethanol baths, clearing with chloroform, and infiltration with paraffin wax as detailed in Table 2.13. Embedded tissue was then stored until required. 7 µm or 20 µm sections were cut using a Leica sledge microtome, transferred into 30 % ethanol solution, floated onto warm water and then picked up on either Vectabond coated slides (Vector) or superfrost slides (BDH). Sections were dried at 37 °C for several hours followed by a 60 °C incubation overnight.

Step	Number of washes	Reagent	Duration (hours)
1	1	70 % ethanol	6
2	2	90 % ethanol	6
3	4	100 % ethanol	6
4	2	Chloroform	6
5	3	Paraffin wax	6

Table 2.13 Paraffin embedding of post mortem brain specimens

2.4.3 Immunohistochemistry

7 µm thick paraffin embedded frontal cortex, cerebellum and hippocampus sections were dewaxed for 3 x 5 minute washes in xylene, followed by rehydration in 100 %, 95 % and 70 % ethanol (Fisher). Slides were incubated in 0.3 % H_2O_2 (Sigma) in methanol (BDH) for 10 minutes to block endogenous peroxidase activity, followed a ddH_2O wash for a minimum of

5 minutes. Sections then underwent antigen retrieval by boiling in citrate buffer in a pressure cooker for 10 minutes at maximum pressure.

Non-specific binding was blocked by incubating slides in 10 % non-fat milk (Marvel) in PBS for 30 minutes at room temperature. Slides were then incubated in the relevant primary antibody (see Table 2.9 above) either at room temperature for 1 hour, or overnight at 4 °C. After two PBS washes, sections were incubated at room temperature for 30 minutes with secondary antibody, followed by two more washes in PBS. The slides were then incubated for 30 minutes with avidin-biotin complex (Dako), washed in ddH_2O, and were then washed twice in PBS. Antibody binding was visualised by the reaction of di-aminobenzenase (DAB, Sigma) chromagen with the peroxidase complex, which results in a dark brown stain. Slides were incubated in 500 µg DAB / 100 mL PBS, activated with 32 µL 30 % H_2O_2 solution for 4 minutes, after which the colour intensity was checked using a light microscope. If a darker stain was required, sections could be replaced into DAB solution for further incubation. Slides were placed into Mayers haemotoxylin (BDH) for 10 seconds to counterstain for cell nuclei, before washing in ddH_2O. Sections were then dehydrated by washing in 70 %, 95 % and 100 % ethanol, cleared in two washes of xylene and were mounted with Depex mounting media (BDH). Slides were then scanned at a 40 x magnification using a Leica SCN400.

2.4.4 Immunofluorescence

In order to facilitate the visualisation of full cell nuclei and nucleoli in human brain, 20 µm thick paraffin embedded frontal cortex sections were used. Sections were first trimmed to approximately 15 mm^2 in size by scraping away excess tissue using a scalpel. Sections were dewaxed and rehydrated as follows: 3 x 5 minute washes in xylene (VWR), 2 x 5 minute washes in 100 % ethanol, 1 x 5 minute wash in 90 % ethanol, 1 x 5 minute wash in 70 % ethanol, 3 x 5 minute washes in ddH_2O.

Following rehydration, both a proteinase antigen retrieval, and a heat-induced antigen retrieval step were carried out, which was found to be necessary to ensure clear visualisation of nucleoli in these sections. The slides were incubated for 2 minutes in 500 µl of pre-prepared proteinase K solution (Dako; diluted in 50 mM Tris-HCl, 15 mM sodium azide, pH 7.5), before washing for 5 minutes in ddH_2O. Samples were lowered into a pressure cooker (Tower 4200H) containing 2 L 10 mM Citrate buffer solution (378 mg/L citric acid; VWR, 2.41 g/L sodium citrate; VWR, in ddH_2O, pH6), which was made by diluting 36 mL citric acid solution (21 g/L citric acid; VWR in ddH_2O), and 164 mL sodium citrate solution (29.41 g/L sodium citrate; VWR in ddH_2O) in 1.8 L ddH_2O. Samples were heated on full power until the pressure indicator was triggered, following which the power was turned down to ⅓ of full power for 4 minutes. The cooker was then taken off the heat for 4 minutes, and then placed under running cold water for 5 minutes. The samples were washed 3 x 5 minutes in ddH_2O, followed by 2 x 5 minute washes in 1 x PBS. Non-specific binding was blocked by incubating slides in 500 µL 10 % FBS in PBS for 30 minutes. Blocking solution was then removed, and slides were incubated in 200 µL of the required primary antibodies diluted in PBS at 4 °C overnight. Samples were then washed 3 x 5 minutes in PBS while rocking, followed by incubation in 200 µL of the required secondary antibodies diluted in PBS, for 1 hour at room temperature. After this incubation the slides were washed for 5 minutes in PBS while rocking, and were then incubated with 500 µL Sudan Black B solution (Sigma; 0.2 % in 70 % Ethanol/ 30 % PBS) for 10 minutes at room temperature to reduce autofluorescence in the tissue. Finally, the slides were washed a further 3 x 5 minutes in PBS while agitating, and the samples were mounted using ProLong® Gold Antifade Mountant with DAPI (Life Technologies) and 24 mm x 50 mm glass coverslips.

2.5 Analysis of immunofluorescence

2.5.1 Confocal Microscopy

Images of immunofluorescent staining were acquired using an LSM710 confocal microscope (Zeiss) using a plan-apochromat 40x/1.4 NA oil immersion objective with Immersol™ Immersion Oil (Zeiss). See Table 2.14 for details of lasers used.

Laser type	Wavelength	Maximum power (mW)
Diode 405-30	405	30
Argon	458, 488, 514	25
DPSS 561-10	561	20
HeNe633	633	5.0

Table 2.14 LSM710 laser details

2.5.1.1 Image capture of RNA foci in SH-SY5Y cells

RNA foci were identified by fluorescence in-situ hybridisation in transfected SH-SY5Y cells (Section 2.2.5). For image capture, fifteen z-stack images were acquired by confocal microscopy consisting of four 2084 x 2084 pixel z-planes each over a 5 µm depth, to give at least 100 transfected cells to analyse. 405 nm, 488 nm and 561 nm lasers all at 2.0 % power were used to excite the samples, with a consistent gain and 90 µm pinhole (2 µm sections, 2.26 AU) used across all images. Z-stacks were converted to a 2D image using a maximum intensity projection preceding image analysis.

2.5.1.2 Image capture of immunofluorescence on human post mortem brain tissue

For each slide of stained human post mortem brain tissue (Section 2.4.4), twenty z-stack images were captured consisting of twelve 2084 x 2084 pixel z-planes each over a 20 µm depth, to give between 200 and 1000 non-inclusion bearing neurons and between 20 and

150 inclusion bearing neurons to analyse in *C9orf72* FTLD patients. 405nm, 488 nm, 561 nm and 633 nm lasers all at 2.0 % power were used to excite the samples, with a 90 µm pinhole (2 µm sections, 2.26 AU) used across all images. Due to the variability of fluorescence intensity between different patient samples, gain was adjusted to peak intensity for each patient.

2.5.2 Image analysis

Analysis of confocal images was performed using Volocity image analysis software (Perkin Elmer).

2.5.2.1 Analysis of RNA foci in SH-SY5Y cells

Analysis protocols were designed to quantify the number of sense RNA foci (see Table 2.15 for a detailed protocol) or antisense RNA foci (see Table 2.16 for a detailed protocol) within the nuclei of cells identified using a co-transfection marker. Data was outputted as the number of foci detected for each cell expressing the co-transfection marker, and exported into Microsoft Excel.

2.5.2.2 Analysis of immunofluorescence on human post mortem brain tissue

Analysis protocols were designed to quantify the 3D volume of the nucleolar markers nucleophosmin (see Table 2.17 for a detailed protocol) or nucleolin (see Table 2.18 for a detailed protocol) within the nuclei of neurons counterstained with NeuN. Both protocols also identified the presence or absence of a poly(GR) inclusion within each neuron. Data was outputted as the volume of nucleolar staining, the volume of DAPI staining and the volume of GR staining for each NeuN positive neuron and for each patient. After exporting the data into Microsoft Excel, neurons for which the program had detected no nucleolar staining or no DAPI staining were excluded from the analysis.

Task	Steps in Volocity protocol
Identify ZsGreen positive cells to give population “ZsGreen”.	Find objects using intensity in 488 channel: threshold 6
	Fill holes in objects
	Exclude objects smaller than 10 μm^2
	Separate touching objects: size guide 800 μm^2
	Exclude objects smaller than 10 μm^2
Identify DAPI stained nuclei to give population “DAPI”.	Find objects using intensity in 405 channel: threshold 20
	Fill holes in objects
	Exclude objects smaller than 10 μm^2
Identify sense foci stained with red FISH probe to give population “RNA foci”	Find objects using intensity in 546 channel: threshold 12
	Exclude objects smaller than 6 μm^2
	Separate touching objects: size guide 0.01 μm^2
	Exclude objects smaller than 0.05 μm^2
Exclude background staining outside of the nucleus to give new population: “RNA foci in DAPI”	Exclude “RNA foci” objects not touching “DAPI” objects
Compartmentalise RNA foci into ZsGreen positive cells	Compartmentalise: Divide “RNA foci in DAPI” between “ZsGreen” objects

Table 2.15 Volocity protocol for quantification of sense RNA foci

Task	Steps in Volocity protocol
Identify DsRed positive cells to give population “DsRed”.	Find objects using intensity in 546 channel: threshold 5
	Fill holes in objects
	Exclude objects smaller than 10 μm^2
	Separate touching objects: size guide 800 μm^2
	Exclude objects smaller than 10 μm^2
Identify DAPI stained nuclei to give population “DAPI”.	Find objects using intensity in 405 channel: threshold 20
	Fill holes in objects
	Exclude objects smaller than 10 μm^2
Identify antisense foci stained with green FISH probe to give population “RNA foci”	Find objects using intensity in 488 channel: threshold 9
	Exclude objects smaller than 6 μm^2
	Separate touching objects: size guide 0.01 μm^2
	Exclude objects smaller than 0.05 μm^2
Exclude background staining outside of the nucleus to give new population: “RNA foci in DAPI”	Exclude “RNA foci” objects not touching DAPI objects
Compartmentalise foci into DsRed positive cells.	Compartmentalise: Divide “RNA foci inside DAPI” between “DsRed” objects

Table 2.16 Volocity protocol for quantification of antisense RNA foci

Task	Steps in Volocity protocol
Identify DAPI stained nuclei to give population “DAPI”	Find objects using standard deviation intensity in 405 channel: threshold 0.8 standard deviations
	Exclude objects smaller than 20 μm^3
	Close: 5 iterations
	Fill holes in objects
	Separate touching objects: size guide 450 μm^3
	Exclude objects smaller than 10 μm^3
Identify nucleophosmin stained nucleoli to give population “nucleoli”	Find objects using standard deviation intensity in 488 channel: threshold 9 standard deviations
	Exclude objects smaller than 0.5 μm^3
	Fill holes in objects
Identify poly(GR) inclusions to give population “GR”	Find objects using intensity in 546 channel: threshold 10
	Exclude objects smaller than 2 μm^3
	Close: 5 iterations
Generate areas 2 pixels dilated from nucleoli to give new population: “Expanded nucleoli” to be used in removal of GR background	Generate a new population by dilating “nucleoli” objects by 2 iterations
Exclude background staining in nucleoli from GR population to give new population: “GR without background”	Subtract “expanded nucleoli” from “GR”
	Exclude objects smaller than 1 μm^3

Identify NeuN positive objects to give population “NeuN”	Find objects using standard deviation intensity in 633 channel: threshold 1.04 standard deviations
	Exclude objects smaller than 1 μm^3
	Close: 5 iterations
	Dilate: 3 iterations
	Fill holes in objects
	Separate touching objects: size guide 500 μm^3
	Exclude objects smaller than 150 μm^3
Define new population “neurons” from DAPI nuclei positive for NeuN staining	Exclude “DAPI” objects not touching “NeuN” objects
	Dilate: 10 iterations
	Separate touching objects: size guide 600 μm^3
	Exclude objects smaller than 300 μm^3
Exclude DAPI not inside Neurons to give new population: “DAPI inside NeuN”	Intersect “DAPI” objects with “neurons” objects
	Exclude objects smaller than 50 μm^3
Compartmentalise nucleoli, DAPI inside NeuN and GR without background into neurons.	Compartmentalise: Divide “nucleoli”, “DAPI inside NeuN” and “GR without background” between neurons.

Table 2.17 Volocity protocol for quantification of nucleophosmin volume in poly(GR) inclusion bearing and non-inclusion bearing neurons

Task	Steps in Volocity protocol
Identify DAPI stained nuclei to give population “DAPI”	Find objects using standard deviation intensity in 405 channel: threshold 0.8 standard deviations
	Exclude objects smaller than 20 μm^3
	Close: 5 iterations
	Fill holes in objects
	Separate touching objects: size guide 450 μm^3
	Exclude objects smaller than 10 μm^3
Identify nucleolin stained nucleoli to give population “nucleoli”	Find objects using standard deviation intensity in 488 channel: threshold 9 standard deviations
	Exclude objects smaller than 0.5 μm^3
	Fill holes in objects
Identify diffuse nucleolin staining to give population “diffuse nucleolin”	Find objects using standard deviation intensity in 488 channel: threshold 5 standard deviations
	Exclude objects smaller than 0.5 μm^3
	Separate touching objects: size guide 80 μm^3
Identify poly(GR) inclusions to give population “GR”.	Find objects using intensity in 546 channel: threshold 10
	Exclude objects smaller than 2 μm^3
	Close: 5 iterations
Generate areas 2 pixels dilated from nucleoli to give new population: “Expanded nucleoli” to be used in removal of GR background	Generate a new population by dilating “nucleoli” objects by 2 iterations

Exclude background staining in nucleoli from GR inclusion objects to give new population: "GR without background"	Subtract "expanded nucleoli" from "GR"
	Exclude objects smaller than 1 μm^3
Identify NeuN positive objects to give population "NeuN"	Find objects using standard deviation intensity in 633 channel: threshold 1.04 standard deviations
	Exclude objects smaller than 1 μm^3
	Close: 5 iterations
	Dilate: 3 iterations
	Fill holes in objects
	Separate touching objects: size guide 500 μm^3
	Exclude objects smaller than 150 μm^3
Define new population "neurons" from DAPI nuclei positive for NeuN staining	Exclude "DAPI" objects not touching "NeuN" objects
	Dilate: 10 iterations
	Separate touching objects: size guide 600 μm^3
	Exclude objects smaller than 300 μm^3
Exclude DAPI not inside neurons to give new population: "DAPI inside NeuN"	Intersect "DAPI" objects with "neurons" objects
	Exclude objects smaller than 50 μm^3
Compartmentalise nucleoli, DAPI inside NeuN and GR without background into neurons.	Compartmentalise: Divide "nucleoli", "diffuse nucleolin", "DAPI inside NeuN" and "GR without background" between "neurons".

Table 2.18 Volocity protocol for quantification of nucleolin volume in poly(GR) inclusion bearing and non-inclusion bearing neurons

2.5.3 Analysis of immunohistochemistry

Analysis of DAB stained post-mortem brain images was carried out using the Definiens Developer XD life portal (Definiens) in collaboration with Matthew Ellis. A ruleset was developed to quantify the proportion of DPR protein inclusion positive cells within the cerebellar granule cell layer and molecular layer of *C9orf72* post-mortem patients (Table 2.19). Four ruleset versions were generated with minor alterations that enabled the analysis of varying DAB staining and hematoxylin signal strengths (Table 2.20, 2.21). Regions to be analysed were determined impartially by selecting a branch residing a pre-specified distance from the edge of the slide. Outlining of the areas for analysis was performed at a low magnification in the absence of the examination of individual cells.

Task	Steps in Definiens protocol
1. Identify region of interest: program differentiates between tissue to be analysed and background around it.	**1a**. Generate low resolution map at 25 % scale (allows image to be processed more quickly when analysing a large image area)
	1b. Find the darkest value in red, blue and green channels for each pixel. Classify the lightest 5th percentile of pixels as background and the darkest 95th percentile as region of interest (ROI).
	1c. Remove isolated objects that comprise less than 5% of total tissue area (removes any extraneous damaged tissue pieces)
2. Manually outline region to be analysed.	**2a**. User manually draws a polygon around the regions of tissue to be analysed in the low resolution map on each slide, and classifies as either “granule cell layer” or “molecular layer”.
3. Removal of very dark debris	**3a**. On the low resolution map, program identifies areas of 500 µm^2 or higher in which the lightest red, green or blue value is under 100 units and removes these from the ROI.

4. Analysis map set up	**4a**. Using the user drawn areas and the region of interest identified, program sets up a separate "Final stain analysis" map for the image analysis to be performed on.	
5. Brown DAB stain identification	**5a**. **Version 1 (Normal DAB staining in granule cell layer).** Program identifies the thresholds for the 95th percentile of brown staining and defines this as "Th_brown", then identifies the top 93rd percentile of brown staining and defines this as "Th_brown_temp" . If Th_brown is below 0.4 units (i.e. the image contains very little brown staining), then this threshold is set to 0.4, and Th_brown_temp is set to 0.35 to avoid identification of false positives.	Different versions of step 5a were used depending on the brain region, darkness of background DAB signal and darkness of hematoxylin stain. Please see tables 2.20 and 2.21 for which versions were used on which specimens.
	5a. **Version 2 (Normal DAB staining in molecular layer).** Program checks the brown value at the 99.9th percentile of brown staining and sequentially checks 0.5 percentile lower until the difference in brown value between two successive percentiles is more than 0.2 units and defines this as the threshold "Th_brown". A second threshold "Th_brown_temp" is set at 1 percentile lower. If Th_brown is lower than 0.4 units (i.e. the image contains very little brown staining), then the program checks the lowest percentile that was reached. If the percentile reached was less than 50.4 (i.e. no inclusions detected in the area), Th_brown is set to 0.6 and Th_brown_temp is set to 0.55. If the lowest percentile reached was above 50.4, Th_brown is set to 0.4 and Th_brown_temp is set to 0.35.	
	5a. **Version 3 (Dark DAB staining in granule cell and molecular layers).** Program identifies the threshold	

	for the 98th percentile of brown staining and defines this as “Th_brown”, then identifies the top 95th percentile of brown staining and defines this as “Th_brown_temp” . If Th_brown is below 0.7 units (i.e. the image contains very few dark aggregates), then this threshold is set to 0.7, and Th_brown_temp is set to 0.65 to avoid identification of false positives.	
	5a. **Version 4 (Very dark DAB identification in granule cell and molecular layers).** Program identifies the threshold for the 98th percentile of brown staining and defines this as “Th_brown”, then identifies the top 95th percentile of brown staining and defines this as “Th_brown_temp” . If Th_brown is below 0.9 units (i.e. the image contains very few dark aggregates), then this threshold is set to 0.9, and Th_brown_temp is set to 0.85 to avoid identification of false positives.	
	5a. **Version 5 (DAB staining against dark hematoxylin in granule cell layer).** Program identifies the threshold for the 95th percentile of brown staining and defines this as “Th_brown”, then identifies the top 93th percentile of brown staining and defines this as “Th_brown_temp” . If Th_brown is below 0.5 units (i.e. the image contains very few dark aggregates), then this threshold is set to 0.5, and Th_brown_temp is set to 0.45 to avoid identification of false positives.	
	5b. Pixels with brown staining higher than the Th_brown threshold are	

	identified by the program. If the identified pixels are bordering with pixels above the Th_brown_temp threshold, the identified area grows out to include them. This process is repeated for 3 iterations.	
	5c. The difference between levels of brown and blue colour is measured, and objects with less than 0.1 units more brown value than blue value are removed. This avoids identification of dark haemotoxylin stain.	
	5d. Brown objects are grown by 1 pixel and shrunk by 1 pixel for 3 iterations. This smooths the object and fills holes.	
	5e. Any single pixel brown objects not touching another brown object are removed.	
6. Blue haemotoxylin stain identification	**6a**. If the mean brown staining value for the image is less than 0.5, i.e. there is low brown background staining, then program checks the blue value at the 99th percentile of blue staining and sequentially checks 1 percentile lower until the difference in blue value between two successive percentiles is more than 0.0025 units and defines this as the threshold "Th_blue". A second threshold "Th_blue_temp" is set at 4 percentile lower. If the mean brown staining value is more than 0.5, i.e there is a high brown background staining, then Th_blue is set to the value of the 40th percentile, and Th_blue_temp is set to the 35th percentile.	
	6b. Pixels with blue staining higher than the Th_blue threshold are identified by the program. If the identified pixels are bordering with pixels above the Th_blue_temp threshold, the identified area grows out to include them. This process is repeated for 3 iterations.	
	6c. Blue objects are grown and shrunk by one pixel for 2 iterations. This smooths the objects and fills holes.	
	6d. Blood vessels are identified by finding	This step was only used in

	blue objects with a length/width ratio of over 3, or a size of over 2000 pixels. These objects are then removed from the nuclei population.	the molecular layer and is absent in the granule cell layer protocol
	6e. Blue objects smaller than 50 pixels are removed.	
7. Separating touching nuclei (summary)	**7a**. Opening: Non-circular objects (with an elliptic fit of < 0.8 where 1 is a perfect circle) are opened using a series of rolling balls of increasing size which erode away the extremities of the cells and accentuate the indentations between cells to facilitate separation.	
	7b. Initial separation: Points within the objects which exhibit a concave angle of > 15 ° are marked as “corners”. A distance map is set up that calculates the distance between all edges in the object. Pairs of corners that define the shortest internal distance in the object are connected to split the cells.	
	7c. Rounding: Merge objects that share 0.5 border distance or more (i.e. one object partially encloses another object) Merge small objects of 50 pixels or less with a larger bordering object when that merge will result in a more circular object.	
	7d. Use of dark stained areas: The 85^{th} percentile is used to define areas of very dark hematoxylin stain within nuclei. These regions undergo 5 iterations of growing by 1 pixel and shrinking by 1 pixel to round the objects and remove holes. These objects are then expanded into the existing nuclei objects, and the border at which these objects meet is used to define a split between the cells.	
	7e. Separating remaining objects over 250 pixels: A distance map is set up for the remaining objects that are larger than 250 pixels. The pixel with the largest distance from the object edge is determined (i.e. the centre point of a nucleus). The distance of pixels from the object edge is then measured in concentric circles expanding out from this centre point. When pixels are found to be further away from the object edge	

	than the previous tier of pixels, this is likely to be a bottleneck between nuclei, and the objects are split at this point. The program undergoes multiple loops of this process until no further objects of over 250 pixels can be divided.
	7f. The areas of brown objects found within blue objects are merged with blue objects to fill holes in nuclei detected due to presence of inclusions.
	7g. Rounding: Objects undergo several iterations of growing by 1 pixel and shrinking by 1 pixel to round the objects and remove holes.
	7h. Remaining small objects are merged into neighbouring nuclei.
8. Compartmentalising brown DAB stained objects into blue nuclei	**8a**. If a nucleus contains more than one brown object, the areas of the inclusions are expanded out into the containing blue object. The blue object is then split along where these domains meet. This results in one cell only containing one inclusion.
	8b. Classify blue objects containing brown objects as positive nuclei. Remaining nuclei are classified as negative nuclei.
	8c. Brown objects not contained within a nucleus are classed as “lone brown”.
9. Export	**9a**. Numbers of “Positive nuclei”, “negative nuclei” and “lone brown” are exported into a Microsoft Excel spreadsheet for analysis.

Table 2.19 Definiens protocol for quantification of the DPR inclusion frequency.

ID	Protocol	Granule Cell Layer	Molecular Layer
A	Normal protocol	Step 5a. Version 1 (Light DAB staining GCL)	Step 5a. Version 2 (Light DAB staining ML) Step 6e. (Blood vessel identification)
B	Dark DAB protocol	Step 5a. Version 3 (Dark DAB staining)	Step 5a. Version 3 (Dark DAB staining) Step 6e. (Blood vessel identification)
C	Very dark DAB protocol	Step 5a. Version 4 (Very dark DAB staining)	Step 5a. Version 4 (Very dark DAB staining) Step 6e. (Blood vessel identification)
D	Dark hematoxylin protocol	Step 5a. Version 5 (DAB detection against dark hematoxylin in GCL)	Step 5a. Version 1 (Light DAB staining ML) Step 6e. (Blood vessel identification)

Table 2.20: Differences between the four rule sets used in Definiens analysis

	Immunostain					
Case no.	**p62**	**GA**	**GP**	**GR**	**AP**	**PR**
9	A	B	A	D	A	A
10	A	C	A	D	A	A
13	A	B	A	D	A	A
14	A	C	A	D	A	A
15	A	C	A	D	A	A
16	A	B	A	D	A	A
17	A	C	A	D	A	A
18	A	B	A	D	A	A
19	A	B	A	D	A	A
20	A	B	A	D	A	A

Table 2.21 Definiens rule sets used for each patient brain specimen

2.6 Statistical analysis

Statistical analysis was carried out using Prism statistical analysis software (GraphPad Software Inc.), with the exception of the nucleolar and nuclear volume comparisons (Chapter 5), which were carried out with the kind help of Dr. Vincent Plagnol using R programming language.

2.7 Software list

Adobe Photoshop

Definiens Developer XD

Microsoft Excel

ND-1000 Nanodrop software (Thermo Scientific)

Peak scanner (Applied Biosystems)

Prism (Graphpad Software Inc.)

Quantity One (BioRad).

Sequence scanner (Applied Biosystems)

Volocity Image analysis software (Perkin Elmer)

Zen black 2011 (Zeiss)

Zen blue 2011 (Zeiss)

Chapter 3: Investigating RNA gain-of-function mechanisms in *C9orf72* FTD/ALS

3.1 Introduction

As detailed in the main introduction (Chapter 1), there are a number of different molecular mechanisms at work that may contribute to the pathology observed in *C9orf72* FTD/ALS patients: loss of C9orf72 protein function, gain of RNA function or gain of DPR protein function. Therefore, an important research question is: what is the impact of each of these different pathways on disease progression? Furthermore, which component in this cascade would be the best target for future treatments? We opted to first investigate the contribution of an RNA gain-of-function mechanism to *C9orf72* FTD/ALS pathology, which has been observed to play an important role in a number of other non-coding repeat pathologies.

It is difficult to determine how different pathological mechanisms might individually lead to neurodegeneration solely by examining patient tissue, due to the vast number of variables interacting in human disease. Instead, we can use model systems to manipulate individual variables in a controlled environment to determine their importance. Cell culture models provide a practical means of rapidly testing a range of different conditions. In order to study the gain-of-function mechanisms of a particular DNA sequence, it can be expressed in a cell line outside of its normal genetic context. This also permits the manipulation of the DNA responsible for disease to determine how it exerts its pathological effect.

We therefore sought to express GGGGCC repeats in isolation from their genomic context in a cell culture system. However, cloning these repeats presents a number of unique difficulties. A common strategy to isolate a DNA sequence from patient tissue would be to

use polymerase chain reaction (PCR) to replicate the sequence. However, this is impeded in *C9orf72* FTD/ALS by the repetitive and 100 percent GC nature of these repeats which results in a high melting temperature and strong secondary structure formation. Successful PCR amplification of the thousands of hexanucleotide repeats observed in patients is therefore unfeasible. In addition, the repeats are highly unstable and frequently undergo deletions when transformed into bacteria for cloning, necessitating the use of specialised cloning protocols.

This chapter details how we overcome these difficulties to generate a range of repeat constructs of different lengths, which allow the study of either RNA gain-of-function in isolation, or RNA and DPR gain-of-functions together. These constructs were then transfected into a SH-SY5Y neuroblastoma cell line, and the formation of both sense and antisense RNA foci were examined in the context of repeat length. These genetic tools were also subcloned into a *Drosophila melanogaster* expression vector and the effect of the different constructs were examined in a fly model of neurodegeneration in collaboration with Professor Linda Partridge and colleagues at the UCL Institute for Healthy Ageing and the Max Planck Institute for Biology of Aging.

3.2 Results

3.2.1 Generation of pure GGGGCC repeat constructs

To generate long tracts of GGGGCC repeat DNA in a plasmid vector, a strategy of recursive directional ligation (RDL) was used to overcome the difficulties imposed by the repetitive, GC-rich nature of this sequence (Figure 3.1). This method builds repetitive sequences of increasing length by seamlessly ligating smaller building blocks of DNA together in a stepwise manner, and has previously been used to synthesise elastin-like polypeptides (Meyer and Chilkoti, 2002). Initial oligonucleotides were designed containing 3 GGGGCC

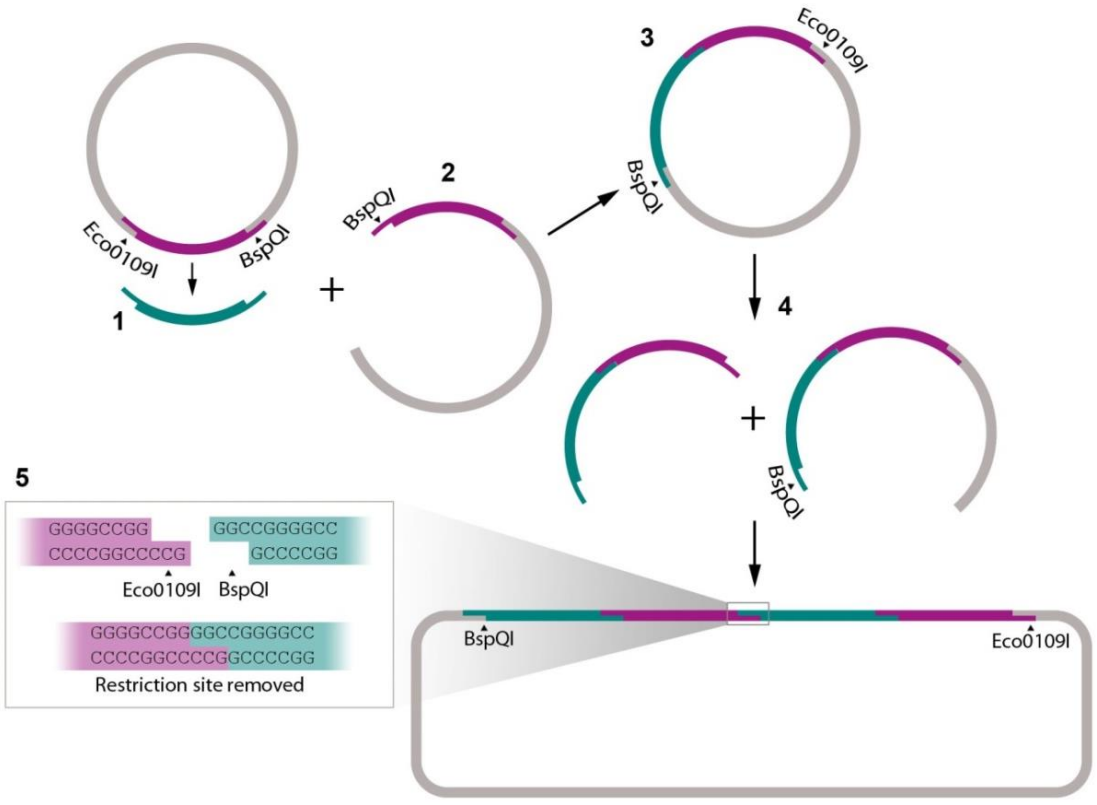


Figure 3.1 Schematic of recursive directional ligation cloning strategy

Recursive directional ligation (RDL) was used to build a range of repeat sizes. (1) A double digest with EcoOI09I and BspQI is carried out on a repeat containing vector to release the repeat insert. (2) The repeat containing vector additionally undergoes a single BspQI digest to open the vector backbone. (3) One or more released repeat inserts are ligated into the vector backbone, generating a product with 2 or more units of the original repeat sequence. (4) Additional rounds of RDL can be carried out to further increase the repeat length exponentially. (5) The ligation is designed so that the restriction site between repeat units is removed with no break in the GGGGCC repeat sequence, allowing the progressive build-up of larger repeats.

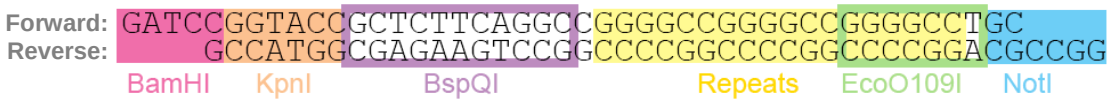


Figure 3.2 Design of pure repeat oligonucleotides

Sequences of the initial forward and reverse oligonucleotides used for generation of pure repeat constructs. Pure repeat oligonucleotides containing 3 repeats were designed with flanking BspQI and EcoOI09I sites to be used in RDL. The oligonucleotide was cloned into empty pBluescript II SK+ using BamHI and NotI restriction sites.

repeats flanked by EcoO109I and BspQI restriction sites (Figure 3.2). Forward and reverse oligonucleotides were annealed and ligated into a pBlueScript II SK+ cloning vector digested with BamHI and NotI, and the ligation product was transformed into chemically competent *E. coli.* For these initial constructs and for other repeat constructs with inserts of 36 repeats or less, colony PCR was carried out to screen for colonies with an insert of the correct size. The repeat containing plasmid DNA was then purified, and two restriction enzyme digestion reactions were set up: 1) a double digest with BspQI and EcoO109I to release the repeat insert, and 2) a single digest with BspQI to open the vector at a site adjacent to the repeats. The digests were resolved on an agarose gel on which the length of the digestion products was checked, after which the repeat insert DNA from the double digest and the linearised vector from the single digest were excised, extracted and ligated together. The overhangs produced by EcoO109I and BspQI enzymes are compatible, and adhesion between the two results in the removal of both restriction sites with no break left in the GGGGCC repeat sequence. The ligation product therefore contains double the number of repeats or more, and restriction sites remain present only at the ends of the total repeat sequence. More than one repeat insert may also ligate into the linearised backbone yielding triple the original number of repeats or more. Indeed, during the first round of RDL with the 3 repeat construct, as many as seven inserts ligated into the vector yielding a construct with 21 total repeats (Table 3.1). Products from the previous ligation can then become the new starting point for the next round of RDL, and can be digested as detailed above to perform additional rounds of RDL. This results in an exponential increase in repeat

Number of repeats	Size (bp)	Generated from	Subcloned into pcDNA3.1+	Subcloned into pUASTattb
3	18	3r oligo	✓	✓
6	36	3 + 3		
9	54	3 + 3 + 3	✓	
12	72	4 x 3		
15	90	5 x 3		
21	126	7 x 3	✓	
36	216	21 + 15	✓	✓
42	252	21 + 21		
63	378	21 +21 + 21	✓*60	
80*	480	63 + 63 deletion product	✓*74	
103*	618	63 + 63 deletion product	✓	✓

Table 3.1 Pure repeat constructs

Pure repeat sizes generated by RDL, and the repeat lengths subcloned into expression vectors. Asterisks denote deletion products.

size with each cycle. Constructs of different sizes can also be ligated to generate numerous different lengths. A number of precautions were taken in conjunction with our recursive directional ligation cloning strategy in an attempt to minimise repeat instability during the cloning process. The *Stbl3 E. coli* strain, which has a mutation in the *recA* recombinase, was used in order to promote repeat stability. It has also previously been shown that passing through stationary growth phase can promote increased deletions in bacteria

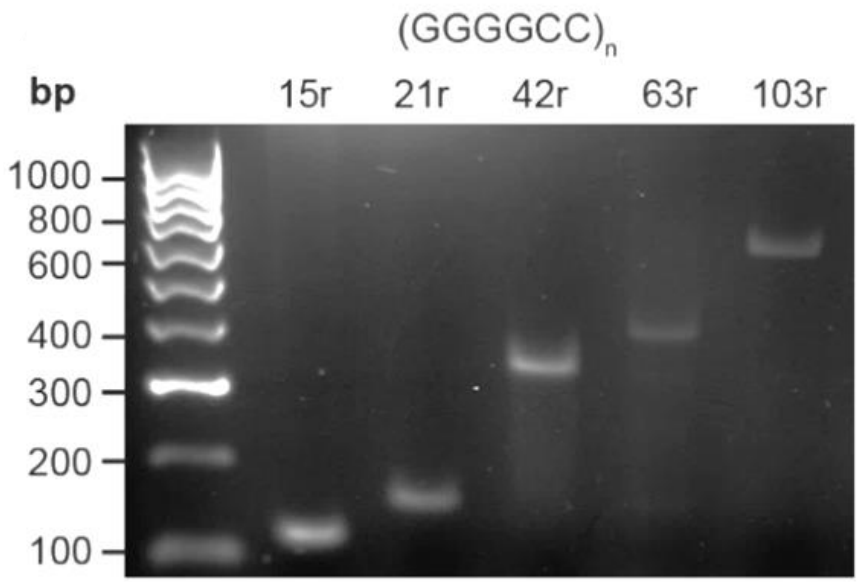


Figure 3.3 Agarose gel displaying pure repeat insert sizes

Pure repeat constructs in pBluescript II SK+ were digested with BspQI and EcoO190I to release repeat fragment for sizing. r = repeats.

(Parniewski et al., 2000), and care was therefore taken to prohibit over-growth of cultures by limiting incubation times and maintaining cultures at 30 °C. Use of these cloning methods allowed the generation of a range of different repeat sizes up to a maximum of 103 pure repeats (103r; Table 3.1; Figure 3.3).

3.2.2 Design of interrupted 'RNA-only' repeat constructs

Pure GGGGCC repeats above 103 repeats in length could not be generated as constructs above this length were highly unstable and underwent frequent deletions when transformed into *E.coli*. To surmount this limit and generate longer repeats such as those seen in human *C9orf72* expansion patients, repeat constructs were designed that contained interruptions: short, non-repetitive segments of DNA interspersed at regular intervals throughout the repeats that act to stabilise the sequence. This strategy has previously been used for several *in vivo* models for myotonic dystrophy without compromising the pathogenicity of the repeats (de Haro et al., 2006; Orengo et al., 2008). A study examining GAA repeats responsible for Friedreich's ataxia found that interruptions

that made up more than 4 % of the total sequence stabilised the repeats, however an interruption content of more than 11 % of the sequence affected DNA secondary structure formation (Sakamoto et al., 2001). We therefore designed the interruption content of our constructs to fall within this range.

An AT-rich six base-pair interruption every 16 GGGGCC repeats (in which the interruption comprised 5.8 % of total sequence) was insufficient in stabilising the repeats in order to achieve more than 160 interrupted repeats, while a six base-pair interruption every 12 repeats (in which the interruption comprised 7.7 % of total sequence) was able to stabilise over 1000 repeats (Figure 3.4). This format of 12 repeats with a 6 nucleotide interruption was therefore chosen to be optimal for our constructs.

After determining that a six base-pair interruption every 12 repeats was optimal for the stabilisation of these repeats, the first reports were published describing RAN translation of the *C9orf72* repeat expansion (Ash et al., 2013; Mori et al., 2013c). We therefore redesigned our interruptions to contain stop codons in all six reading frames in order to generate 'RNA-only' constructs for the purpose of dissecting RNA and DPR protein gain-of-function disease mechanisms.

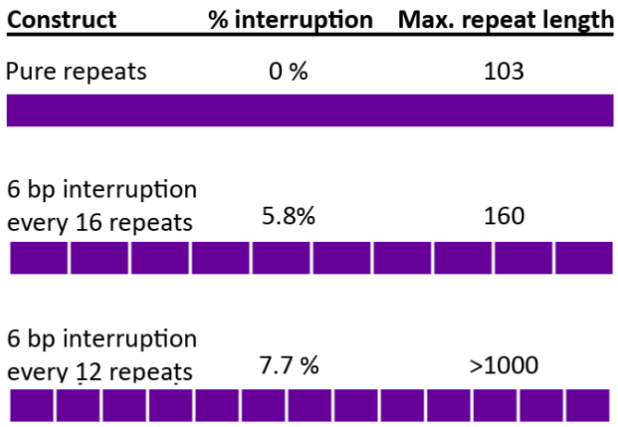


Figure 3.4 Properties of pure and interrupted repeat constructs.

Schematic diagram of the proportion of sequence comprised of interruptions. Purple bars represent GGGGCC repeat DNA. White bars represent 6 base pair AT rich interruptions

Three different AT-rich interruption sequences of 6 nucleotides in length were designed, which we termed interruptions A, B and C. Each interruption contained stop codons in two of the 6 reading frames (one sense frame and one antisense frame) to prohibit peptide production in all frames when expressed together (Figure 3.5). Interruptions were designed to occur every 12 repeats, alternating between the three different sequences, with a stop codon therefore occurring in each individual reading frame every 36 repeats. Three alternating interruption sequences were required as it was not possible to introduce more than two stop codons into a 6 base-pair AT rich sequence without either a) introducing an ATG start codon or b) introducing an EcoO109I restriction site.

In order to insert the interruptions into the repeat sequence, three pairs of oligonucleotides were designed flanked by BspQI and EcoO109I overhangs, each containing one of the 3 interruptions plus 4 hexanucleotide repeats (Figure 3.6). Each pair of oligonucleotides were annealed and ligated into a BspQI and EcoO109I digested pBlueScript II SK+ vector containing 8 pure repeats. This resulted in three constructs that contained 12 repeats plus either interruption A (12ri A), B (12ri B) or C (12ri C). 12ri A was then singly digested with BspQI to open the plasmid adjacent to the repeat insert, while

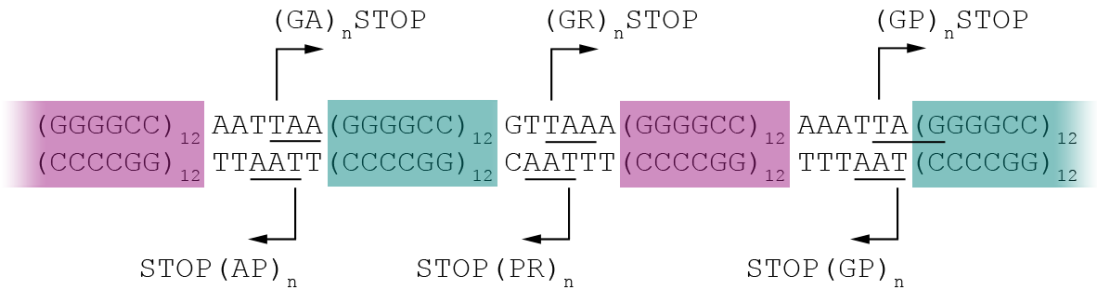


Figure 3.5 Design of stop codon containing interruptions

In the RNA-only constructs, three AT rich interruptions occur every 12 GGGGCC repeats, each containing a stop codon in one sense and one antisense frame. This schematic shows the frames of RAN translation that are stopped in each interruption, resulting in one stop codon occurring in each frame every 36 repeats.

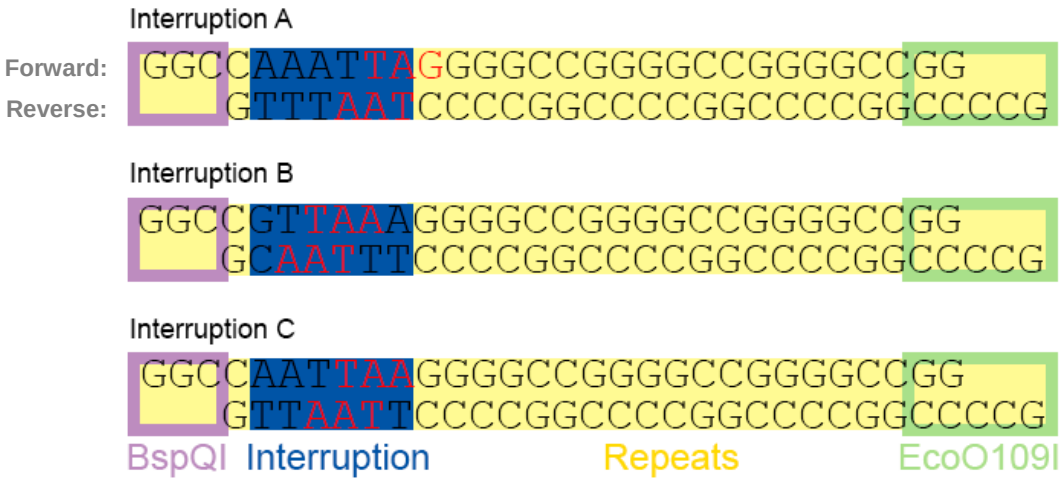


Figure 3.6 Design of Interrupted repeat oligonucleotides

Sequences of the initial forward and reverse oligonucleotides used for generation of pure repeat constructs. Three different forward and three different reverse interrupted oligonucleotides were designed, each containing a different interruption with stop codons in 2 frames and 4 GGGGCC repeats, which were cloned into pure 8 repeat constructs in pBluescript II SK+ to give a total of 12 repeats and an interruption.

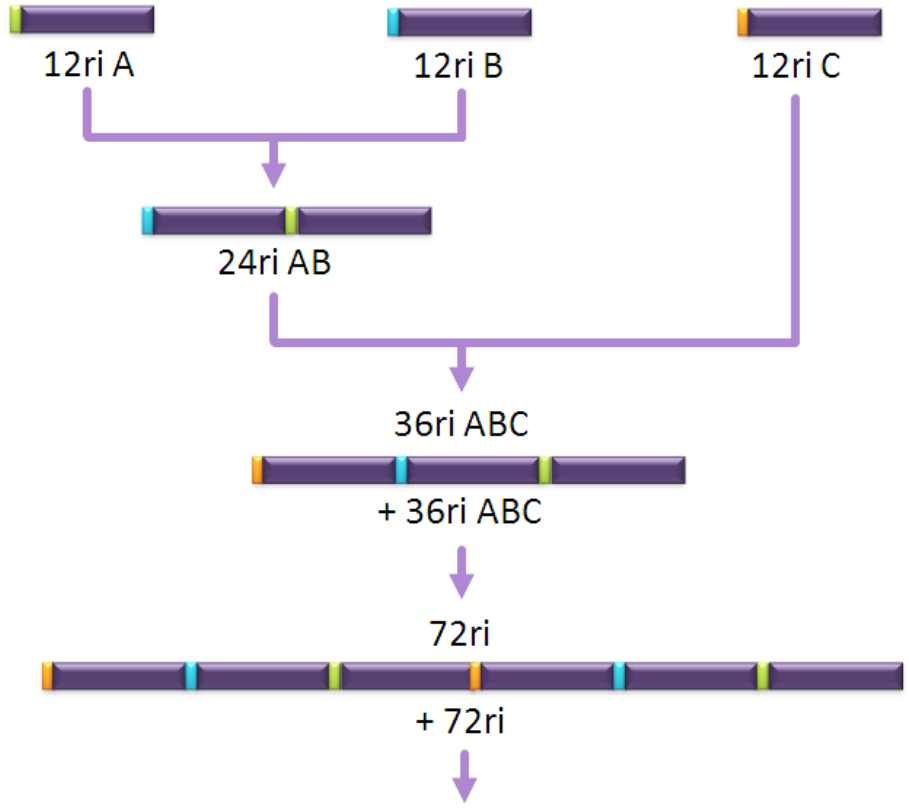


Figure 3.7 Generation of alternating interruption constructs

Flow chart detailing the initial cloning steps taken to generate the initial RNA-only constructs with three distinct alternating interruptions. Purple bar represents GGGGCC repeat sequence. Green, blue and orange bars represent interruptions A, B and C. Arrows illustrate the ligation order.

Number of repeats	Units of 36	No. of interruptions	Size of insert (bp)	Generated from	Subcloned into pcDNA3.1+	Subcloned into pUASTattB
12	0	1	78	8 + 4r with interruption oligo		
24	0	2	156	12 + 12		
36	1	3	234	24 + 12	✓	✓
72	2	6	468	36 + 36		
108	3	9	702	72 + 36	✓*107	✓
144	4	12	936	72 + 72		
288	8	24	1,872	144 + 144	✓	✓
576	16	48	3,744	288 + 288	✓	✓
864	24	72	5,616	288 + 288 +288		
1152	32	96	7,488	864 + 288	✓	✓

Table 3.2 RNA-only interrupted repeat constructs

Interrupted repeat sizes generated by RDL, and the repeat lengths subcloned into expression vectors. Asterisks denote deletion products

12ri B underwent double digestion to release the insert, and these two products were ligated to yield 24 repeats containing interruptions A and B (24ri AB; Figure 3.7). Finally, 24ri AB was singly digested with BspQI to open the plasmid adjacent to the repeat insert, while 12ri C underwent double digestion to release the insert, and these two products were ligated to yield 36 repeats containing interruption A, B, and C (36ri ABC; Figure 3.7). This 36ri ABC construct was then used as the initial building block for the process of RDL,

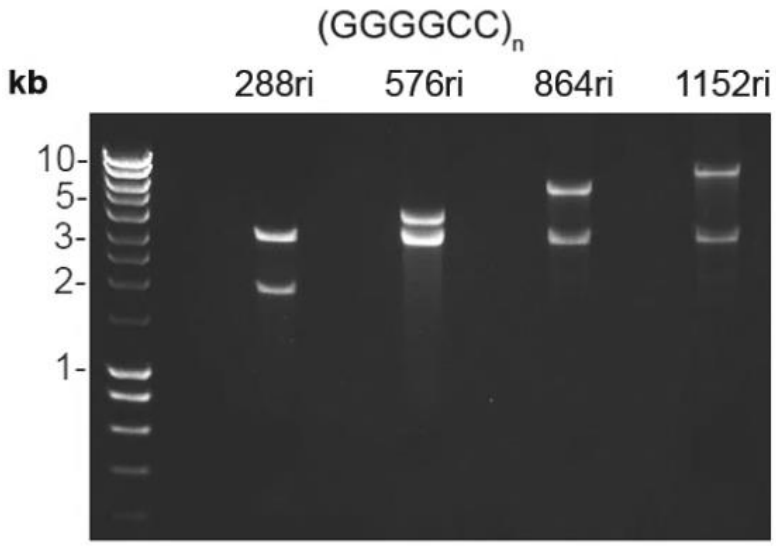


Figure 3.8 Agarose gel displaying interrupted repeat insert and cloning vector sizes

Larger interrupted constructs in pBluescript II SK+ (vector band at 3 kb) were digested with BamHI and NotI to release repeat fragment for sizing. ri = interrupted repeats.

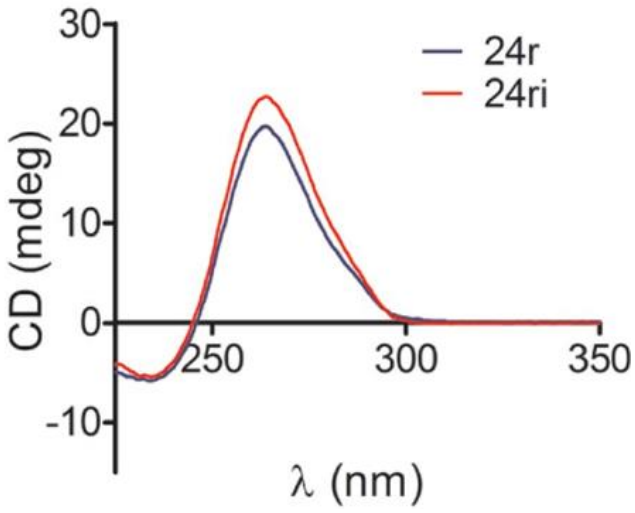


Figure 3.9 G-quadruplex formation by pure and interrupted repeats

Circular dichroism (CD) spectra for pure and interrupted repeats. Both 24 pure repeats (24r) and 24 repeat constructs interrupted every 12 repeats (24ri) exhibit CD spectra characteristic of G-quadruplex formation, indicating the retention of secondary structure formation by interrupted repeats.

which proceeded as described in section 3.2.1. A range of RNA-only construct sizes were built up to a largest stable size of 1152 interrupted repeats (1152ri) requiring 5 RDL cycles from the 36 interrupted repeat block (Table 3.2; Figure 3.8). Circular dichroism carried out by Dr. Andrew Nicoll on RNA from pure repeats and 24 repeat constructs interrupted every 12 repeats showed that both structures formed spectra characteristic of G-quadruplexes (Figure 3.9), indicating that these repeats retain their ability to form disease relevant secondary structure.

3.2.3 Subcloning of repeat constructs into mammalian and fly expression vectors

A range of different sized pure repeats (r) and interrupted repeats (ri) were subcloned into a pcDNA3.1+ vector for expression in mammalian cells, and also into pUASTattB for expression in *Drosophila melanogaster* (Figure 3.10). For subcloning into pcDNA3.1+ the vector and each repeat construct to be subcloned was digested with BamHI and NotI, while for subcloning into pUASTattB, the vector and repeat constructs to be cloned were digested with EcoRI and NotI. The digests were resolved on an agarose gel to check the size

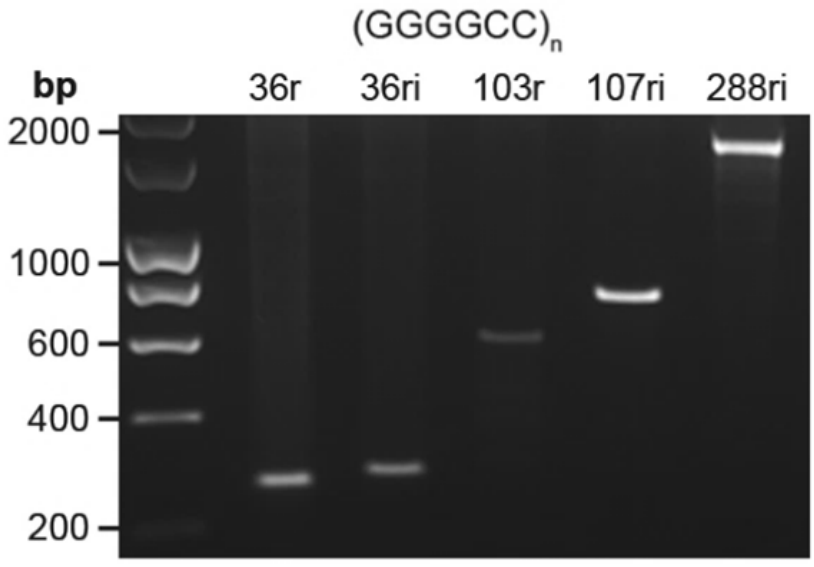


Figure 3.10 Agarose gel displaying subcloned repeat insert sizes

Pure and interrupted repeat constructs cloned into pUASTattb were digested with EcoRI and NotI to release repeat fragment for sizing. r = pure repeats, ri = interrupted repeats.

of the digestion products against the expected size using the molecular weight markers. Insert bands of the correct size that did not show deletion products or smearing were extracted and ligated into the desired vector, and high concentration DNA at transfection grade purity was prepared by maxiprep.

3.2.4 Sequencing of repeat constructs

To check for deletions and mutations and to determine the exact repeat length, sequencing was attempted for all constructs generated. Due to the repetitive, GC-rich nature of the constructs, the maximum lengths for which sequencing was possible in full were 103r and 108ri. Both ends of the 288ri, 576ri and 1152ri repeats were sequenced as far as possible to screen for mutations, but exact size could not be determined by sequencing. The majority of constructs were found to be free from mutations, and those constructs found to contain point mutations were discarded.

3.2.5 Optimisation of SH-SY5Y transfection

To characterise the different constructs generated, transient transfection was carried out in the neuroblastoma cell line SH-SY5Y. A neuronal-like cell line was chosen as neurons are the cells principally affected by degeneration in *C9orf72* FTD/ALS. In order to distinguish which cells had been transfected with the repeat constructs, cultures were co-transfected with either the green fluorescent marker ZsGreen or the red fluorescent marker DsRed, with a 10:1 ratio of repeat construct to co-transfection marker. To optimise the Lipofectamine2000 transfection protocol, a range of different cell densities, DNA quantities, transfection reagent quantities and transfection reagent incubation times were tested. Cells were examined by fluorescence microscopy, with transfection conditions chosen which: 1) maximised the percentage of transfected cells, which was determined by the percentage of ZsGreen-expressing cells and 2) minimised the loss of cell viability caused by the transfection process, which was approximated by the frequency of condensed or

blebbing DAPI stained nuclei. The chosen transfection protocol resulted in ZsGreen expression in approximately 25% of cells. To test for the co-occurrence of a transfected plasmid with a co-transfected marker, DsRed and ZsGreen were transfected into cells at a 10:1 ratio of DsRed to ZsGreen. This resulted in a 95 % co-occurrence between the two markers, indicating that the co-transfection marker can signify transfected cells with reasonable reliability (Figure 3.11).

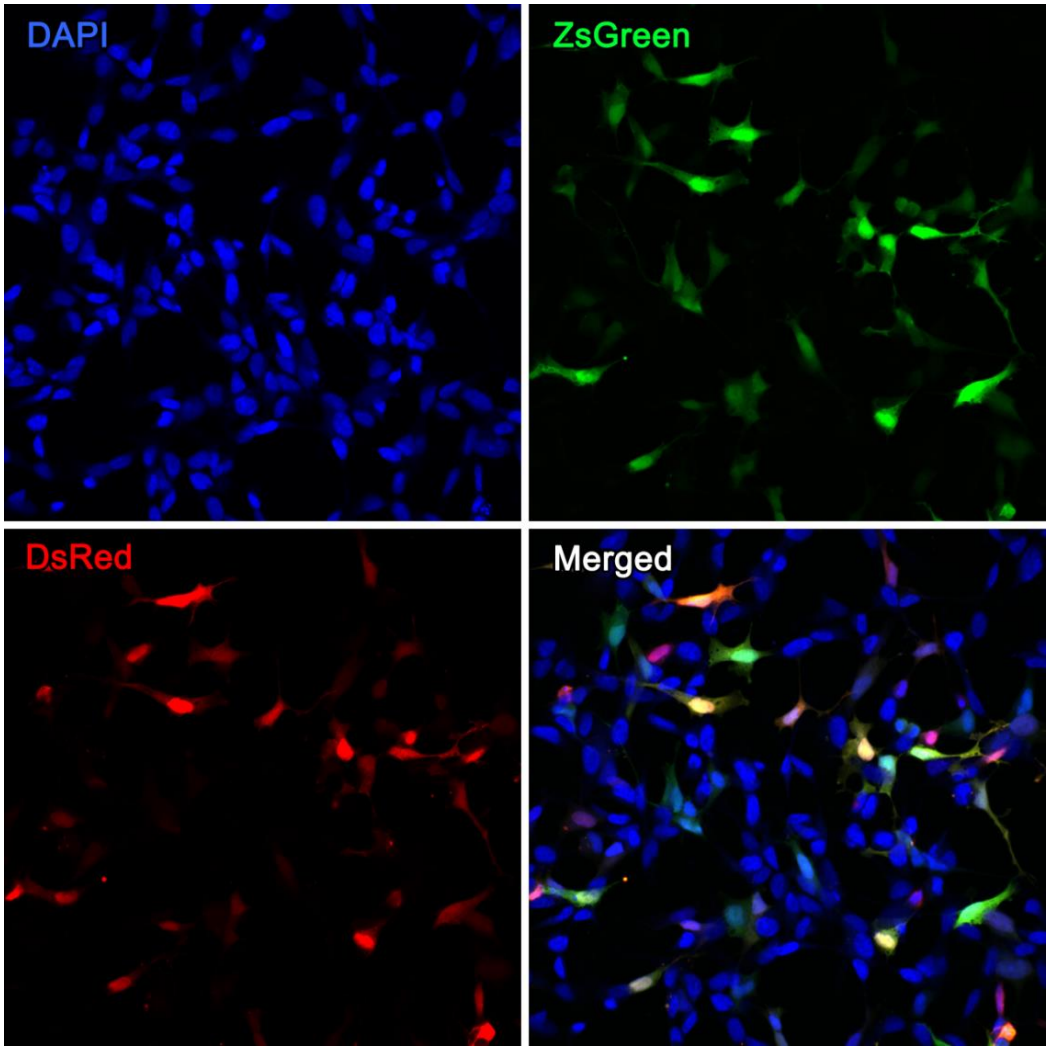


Figure 3.11 Co-occurrence of DsRed and ZsGreen in optimised SH-SY5Y transfection protocol

SH-SY5Y cells co-transfected with ZsGreen (green) and DsRed (red). DAPI nuclear stain shown in blue.

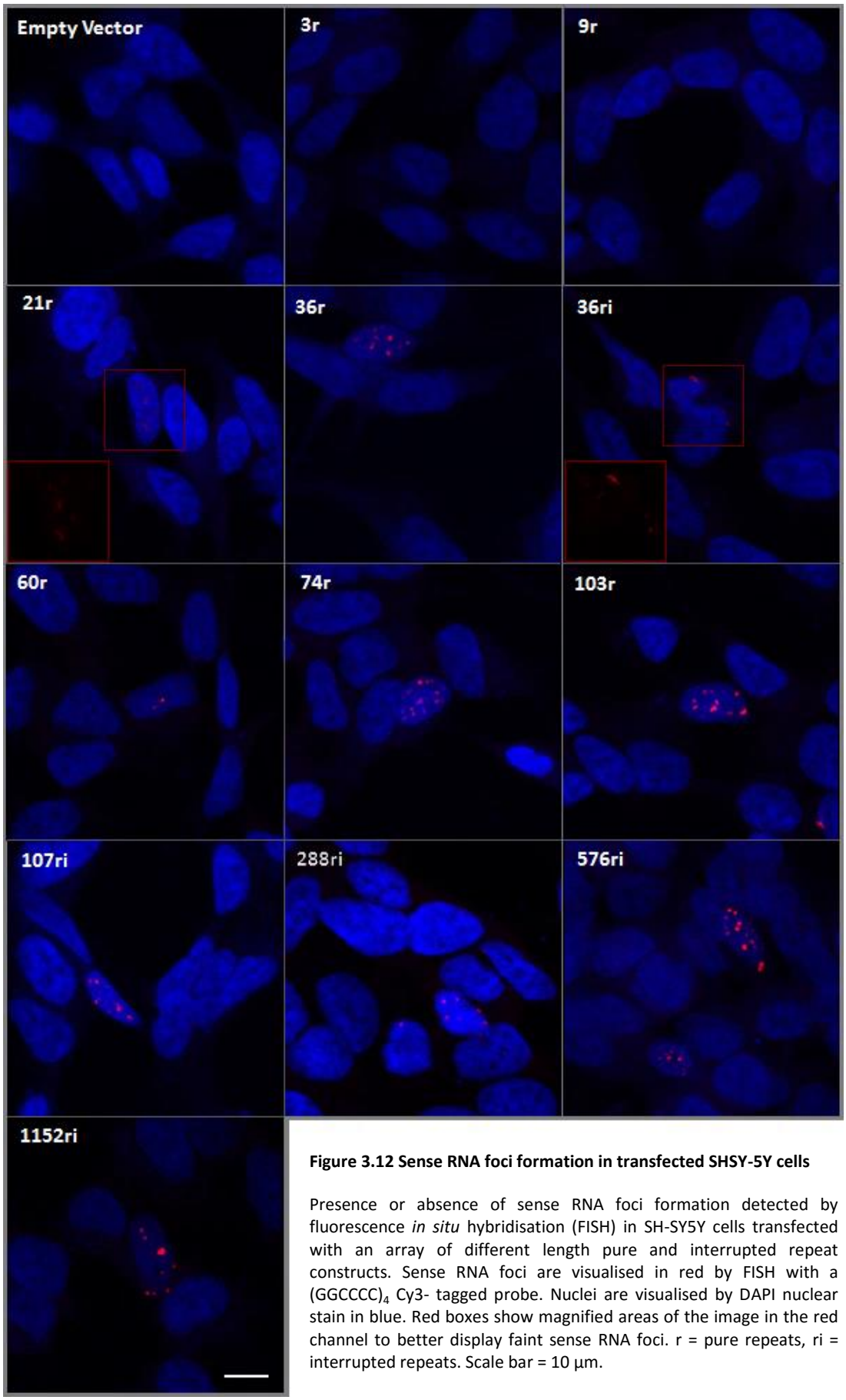


Figure 3.12 Sense RNA foci formation in transfected SHSY-5Y cells

Presence or absence of sense RNA foci formation detected by fluorescence *in situ* hybridisation (FISH) in SH-SY5Y cells transfected with an array of different length pure and interrupted repeat constructs. Sense RNA foci are visualised in red by FISH with a $(GGCCCC)_4$ Cy3- tagged probe. Nuclei are visualised by DAPI nuclear stain in blue. Red boxes show magnified areas of the image in the red channel to better display faint sense RNA foci. r = pure repeats, ri = interrupted repeats. Scale bar = 10 µm.

3.2.6 Characterisation of sense RNA foci formation

In order to study sense RNA foci formation by repeat constructs in cells, a range of pure and interrupted 'RNA-only' repeat constructs of different lengths were transfected into SH-SY5Y cells, and RNA fluorescence *in situ* hybridisation (FISH) was carried out with a Cy3-labelled $(GGCCCC)_4$ RNA probe, complementary to the sense GGGGCC RNA. Subsets of cells transfected with pure or interrupted repeat constructs containing 21 repeats or more were observed to exhibit distinct sense RNA foci formation when examined by confocal fluorescence microscopy, which was not observed in empty vector, 3r or 9r transfected cells (Figure 3.12). To quantify the number of sense foci present in each ZsGreen-positive cell, images of transfected cells were acquired by confocal microscopy, and an automated image analysis protocol was developed using Volocity image analysis software. DAPI stained nuclei were identified in the blue 405 nm channel, cells expressing the ZsGreen co-tranfection marker were identified in the green 488 nm channel, and RNA foci identified by the FISH probe were identified in the red 546 nm channel. RNA foci identified outside of DAPI stained areas were excluded by the software due to high background fluorescence in the cytoplasm that lead to a higher false positive rate than was acceptable. Therefore, only nuclear RNA foci were counted.

The number of RNA foci present within each ZsGreen-positive cell was quantified for each coverslip of transfected cells. These values were used to determine the percentage of ZsGreen-positive cells that contained one or more RNA foci and the mean number of RNA foci present in positive cells. In SH-SY5Y cells transfected with pure repeat constructs, a significant positive correlation was observed between repeat length and the percentage of ZsGreen-positive cells exhibiting sense RNA foci when a linear regression was performed ($p<0.0001$, $r^2 = 0.6194$; Figure 3.13a). Repeat length in interrupted constructs was also positively correlated with the percentage of ZsGreen-positive cells exhibiting sense RNA foci, with some saturation in the percentage of foci positive cells observed in the largest

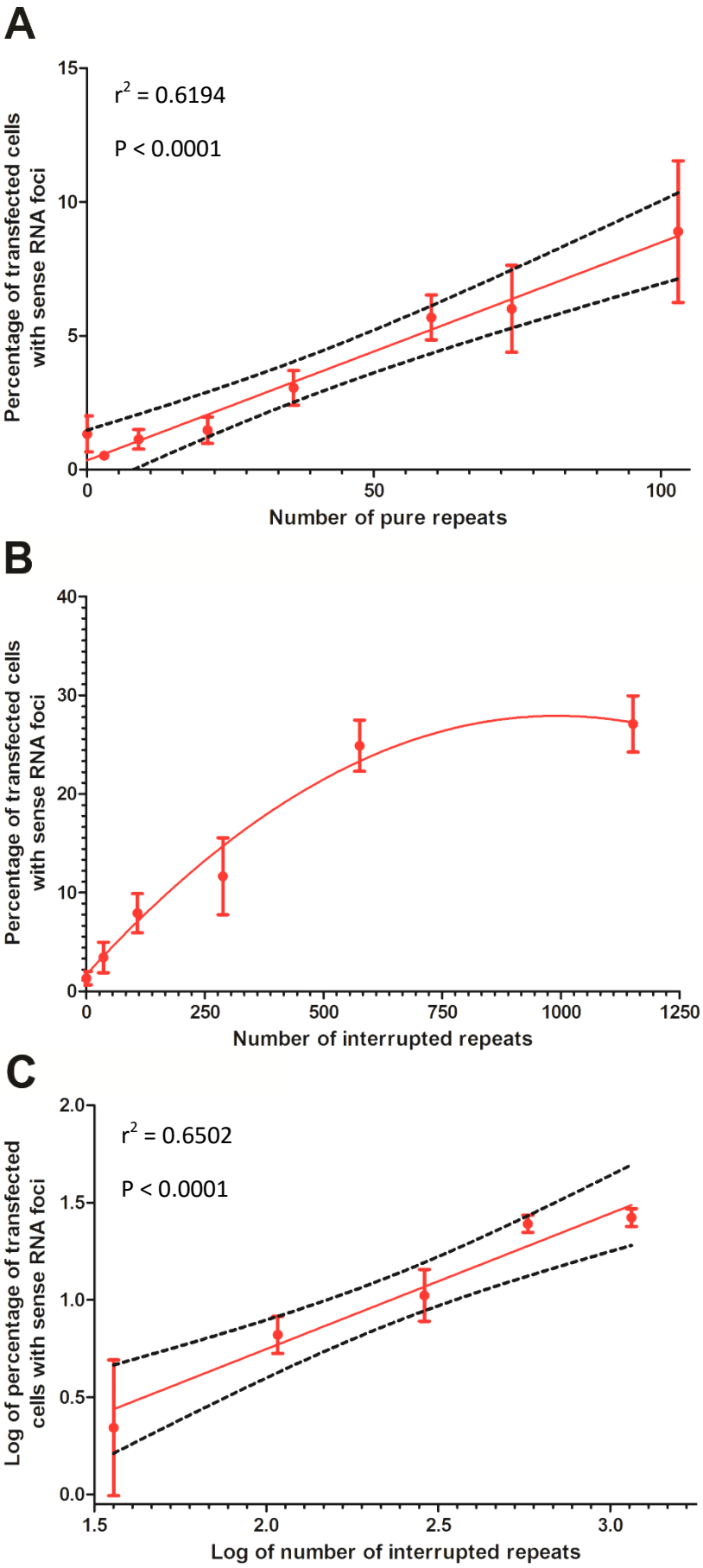


Figure 3.13 Correlation of sense RNA foci formation with repeat length in SH-SY5Y cells

A)The percentage of transfected cells with sense RNA foci was plotted against repeat length for pure repeats and a linear regression was performed
B) The percentage of transfected cells with sense RNA foci was plotted against repeat length for interrupted repeats.
C) As the percentage of transfected cells with sense RNA foci appeared to reach saturation in larger interrupted repeats, log of number of repeats was plotted against log of percentage of transfected cells with sense RNA foci. A linear regression was performed on this transformed interrupted repeat data.
R^2 values are shown as a measure of goodness of fit. P values shown denote the significance of a linear trend with a non-zero slope. Error bars represent standard error of mean (SEM). Red lines indicate line of best fit. Dashed black lines represent 95% confidence band for the line of best fit. A minimum of 3 experimental replicates were carried out for each repeat length.

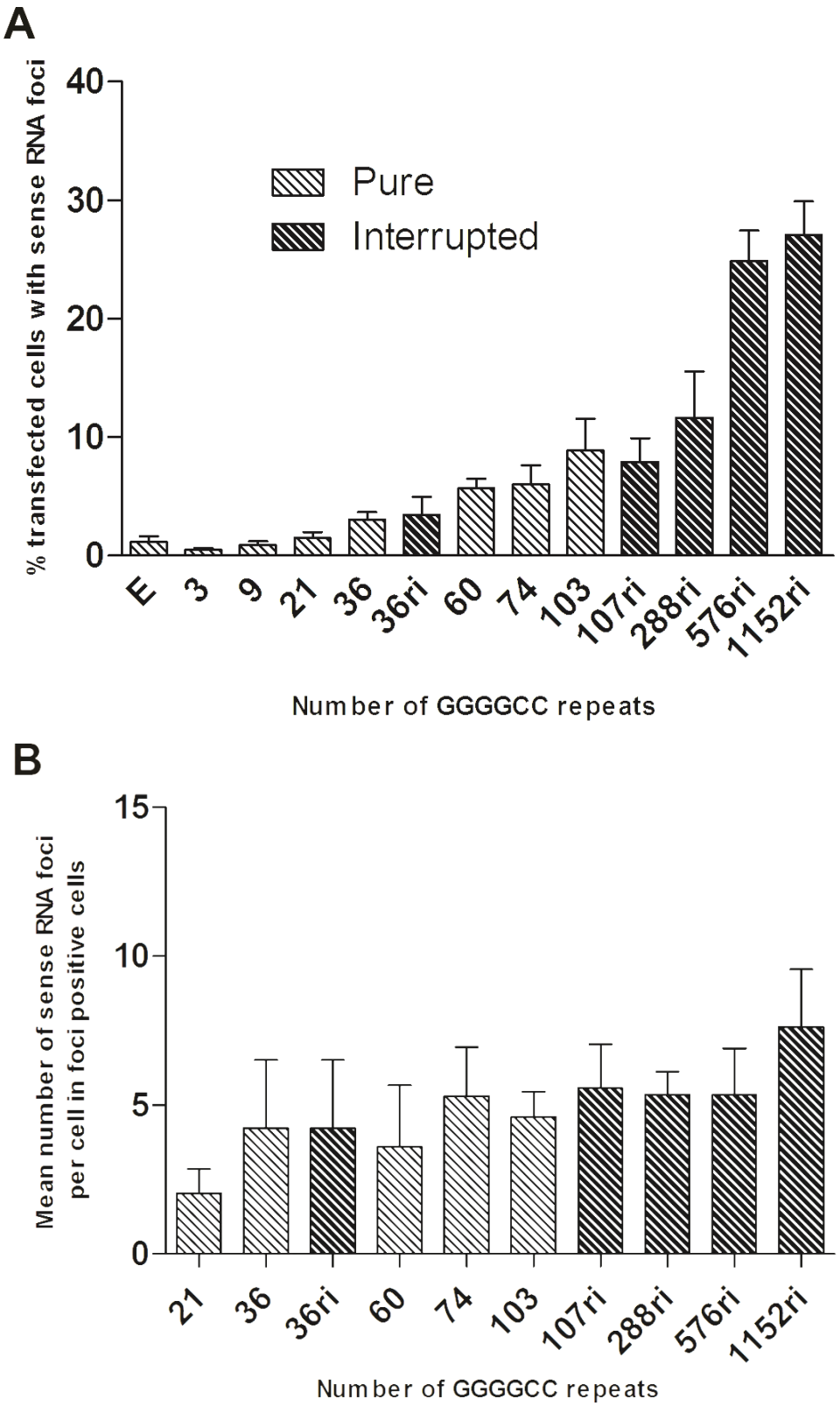


Figure 3.14 Quantification of sense RNA foci formation in SH-SY5Y cells

A) The percentage of transfected cells containing sense RNA foci in SH-SY5Y cells transfected with an array of different length pure (light bars) and interrupted (dark bars) repeat constructs. B) Mean number of sense RNA foci per cell exhibited by RNA-foci-positive SH-SY5Y cells transfected with an array of different length pure and interrupted repeat constructs. Error bars represent standard error of mean (SEM). A minimum of 3 experimental replicates were carried out for each repeat length. r = pure repeats, ri = interrupted repeats.

repeat lengths (Figure 3.13b). A log transformation was carried out on both x and y values for these data to linearise the observed trend, and a linear regression showed a significant positive correlation between the percentage of sense RNA foci positive cells and interrupted repeat length ($p<0.0001$, $r^2= 0.6502$; Figure 3.13c). Bar graphs were plotted for cells transfected with each construct to compare sense RNA foci formation by cells transfected with pure and interrupted constructs (Figure 3.14). The presence of interruptions in RNA-only constructs had no apparent effect on RNA foci formation: the same proportion of cells exhibited foci in 103r (8.9 ± 2.6 %) and 107ri (7.9 ± 2.0 %), or 36r (3.1 ± 0.64 %) and 36ri (3.4 ± 1.5 %) transfected cells, with no significant difference found between the two pairs by one-way analysis of variance (ANOVA) with Bonferroni post-hoc test on the selected columns. Sense RNA foci were exhibited most frequently in cells transfected with our largest repeat construct 1152ri, with foci observed in 27 ± 2.9 % of transfected cells (SEM; Figure 3.14a). There was a small background level of RNA foci-positive cells picked up in empty pcDNA3.1+ vector, 3r and 9r transfected cells, which may be due to the automated RNA foci counting process and the occasional presence of autofluorescent debris among cells. While RNA foci are formed in only a low percentage of cells in the 36r and 36ri constructs, this does not prohibit the occasional formation of larger numbers of foci in a single cell: several 36r and 36ri transfected cells exhibited over 20 foci, with a highest foci burden in a single cell of 43 foci. Quantification of RNA foci number per cell in foci-positive cells reveals no clear correlation between average number of foci in foci-positive cells and repeat length (Figure 3.14b). To further examine whether any differences in the distribution of RNA foci number per cell could be seen between constructs, histograms were plotted for the percentage of sense RNA foci positive cells that formed a given number of foci (Figure 3.15). The number of sense RNA foci exhibited by transfected cells was binned into groups of 5, and the percentage of the sense RNA foci positive cells exhibiting a number of foci within each bin was plotted. The proportion of the

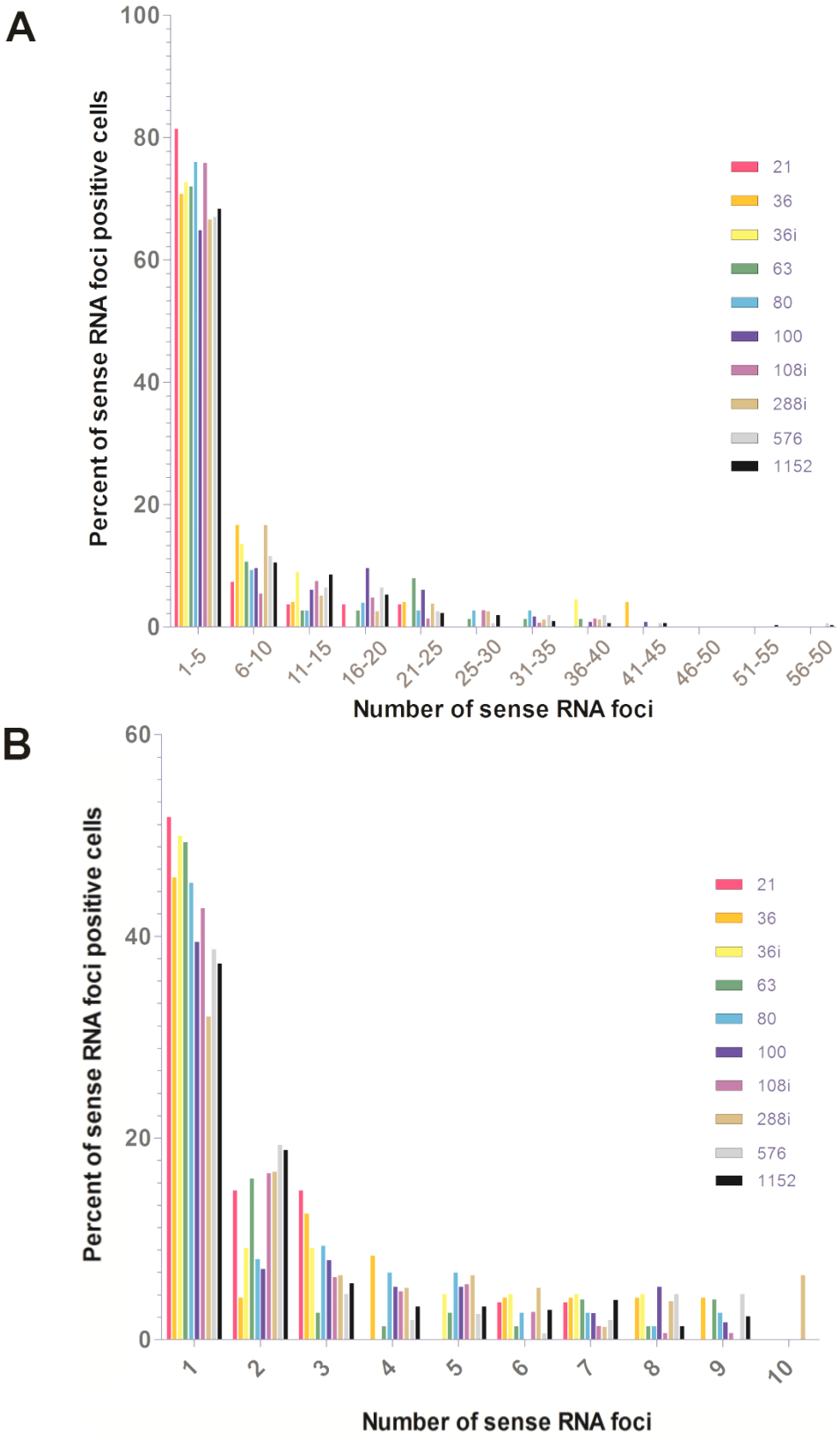


Figure 3.15 Frequency distribution of number of sense RNA foci per cell

Histogram displaying the proportion of cells exhibiting sense RNA foci in SH-SY5Y cells transfected with different repeat constructs. A) The number of sense RNA foci expressed in all sense RNA foci positive cells was binned into groups of 5, with each bar representing the percentage of cells transfected with a given repeat construct that express a quantity of sense RNA foci within that bin range. B) The percentage of sense RNA foci positive cells exhibiting between 1 and 10 sense RNA foci was examined in more detail.

population exhibiting a given number of sense RNA foci was observed to diminish with increasing foci number (Figure 3.15a). As approximately 80 % of foci positive cells exhibited between 1 and 10 foci, a second histogram was plotted for this range in which cells exhibiting individual numbers of foci was analysed, and around half of all sense RNA foci positive cells were observed to exhibit one RNA focus (Figure 3.15b). No clear difference in the distribution of RNA foci number between different repeat lengths was apparent.

3.2.7 Characterisation of antisense RNA foci formation

A smaller subset of the constructs generated were co-transfected into SH-SY5Y cells with a DsRed transfection marker for examination of antisense RNA foci formation. RNA FISH was carried out using an Alexa 488-labelled $(GGGGCC)_4$ RNA probe complementary to the GGCCCC antisense repeat transcript. A subset of cells transfected with pure or interrupted repeat constructs containing 103 repeats or more were observed to exhibit antisense RNA foci formation when examined by confocal fluorescence microscopy (Figure 3.16), demonstrating that these can be formed in cultured cells despite a lack of a promoter in the antisense direction. A second protocol on Volocity image analysis software was developed as described above to detect antisense foci formation, in which RNA foci were located in the green 488 nm channel, and the DsRed co-transfected cells in the red 546 nm channel. The percentage of transfected cells expressing antisense RNA foci correlated positively with repeat length as seen for sense RNA foci formation, with 1152ri again forming RNA foci in the greatest percentage of cells at 22 ± 2.5 % (Figure 3.17a). No clear antisense RNA foci formation was observed in 36r and 36ri transfected cells, while approximately 1 % of 103r and 107ri transfected cells formed antisense foci, compared to the 8-9 % of 103r and 107ri transfected cells which were positive for sense foci (Figure 3.18a). In addition, no more than 2 antisense RNA foci per cell were detected in all 103r and 107ri transfected cells (Figure 3.17b). Antisense foci were formed in around 20 % fewer

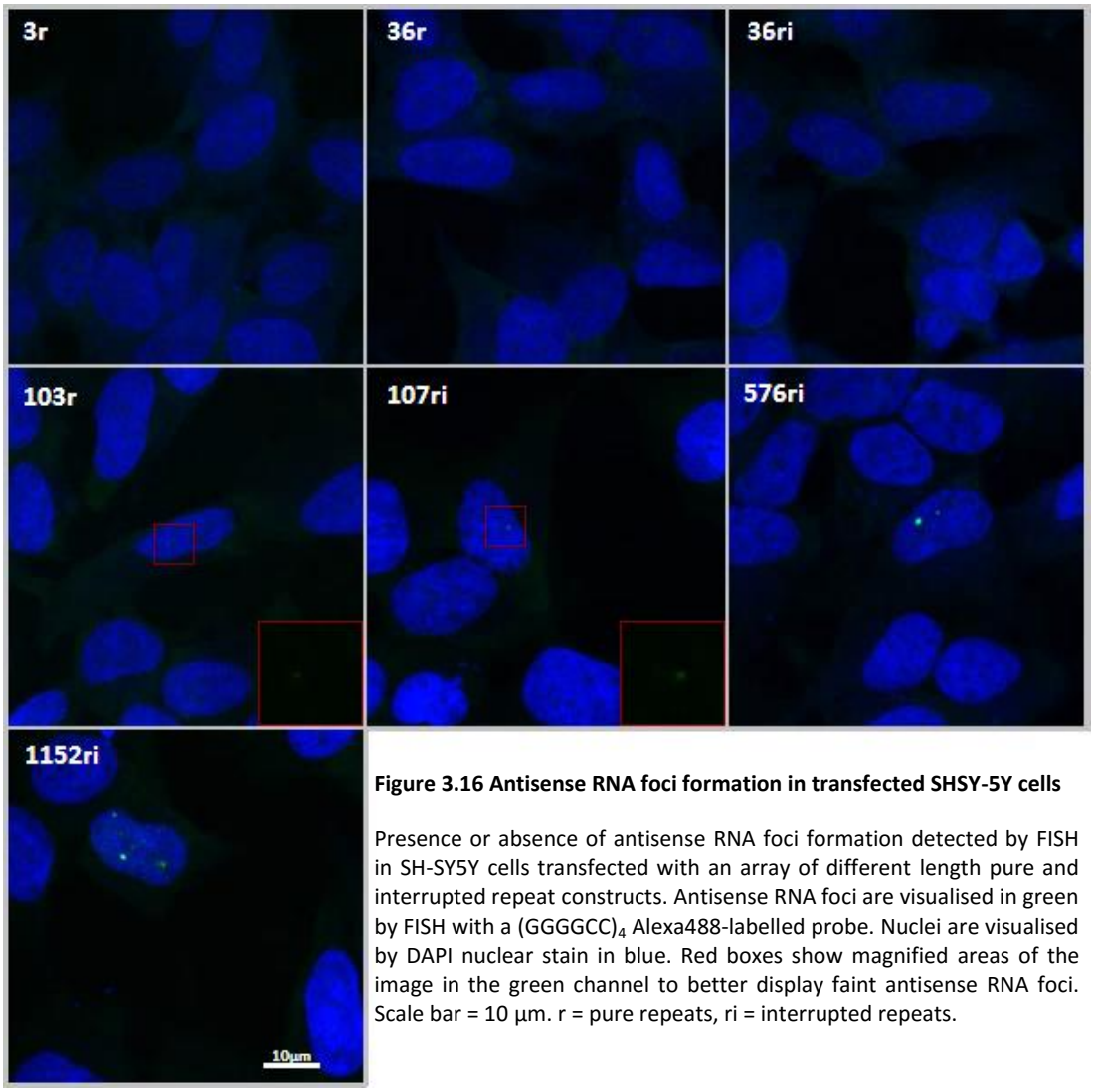


Figure 3.16 Antisense RNA foci formation in transfected SHSY-5Y cells

Presence or absence of antisense RNA foci formation detected by FISH in SH-SY5Y cells transfected with an array of different length pure and interrupted repeat constructs. Antisense RNA foci are visualised in green by FISH with a $(GGGGCC)_4$ Alexa488-labelled probe. Nuclei are visualised by DAPI nuclear stain in blue. Red boxes show magnified areas of the image in the green channel to better display faint antisense RNA foci. Scale bar = 10 µm. r = pure repeats, ri = interrupted repeats.

cells in 576ri and 1152ri transfected samples compared to sense foci, with the average number of foci formed per cell in foci-positive cells also reduced by over half (Figure 3.18). Histograms were plotted for the number of antisense foci present in each foci-positive cell, with the number of foci binned into groups of 5 (Figure 3.19). The percentage of antisense RNA foci positive cells that formed a given number of antisense RNA foci was plotted. As was seen for sense foci, the proportion of the population exhibiting a given number of antisense RNA foci was observed to diminish with increasing foci number (Figure 3.19a). As approximately 90 % of antisense RNA foci positive cells exhibited between 1 and 10 foci, a second histogram was again plotted for this range in which cells exhibiting individual

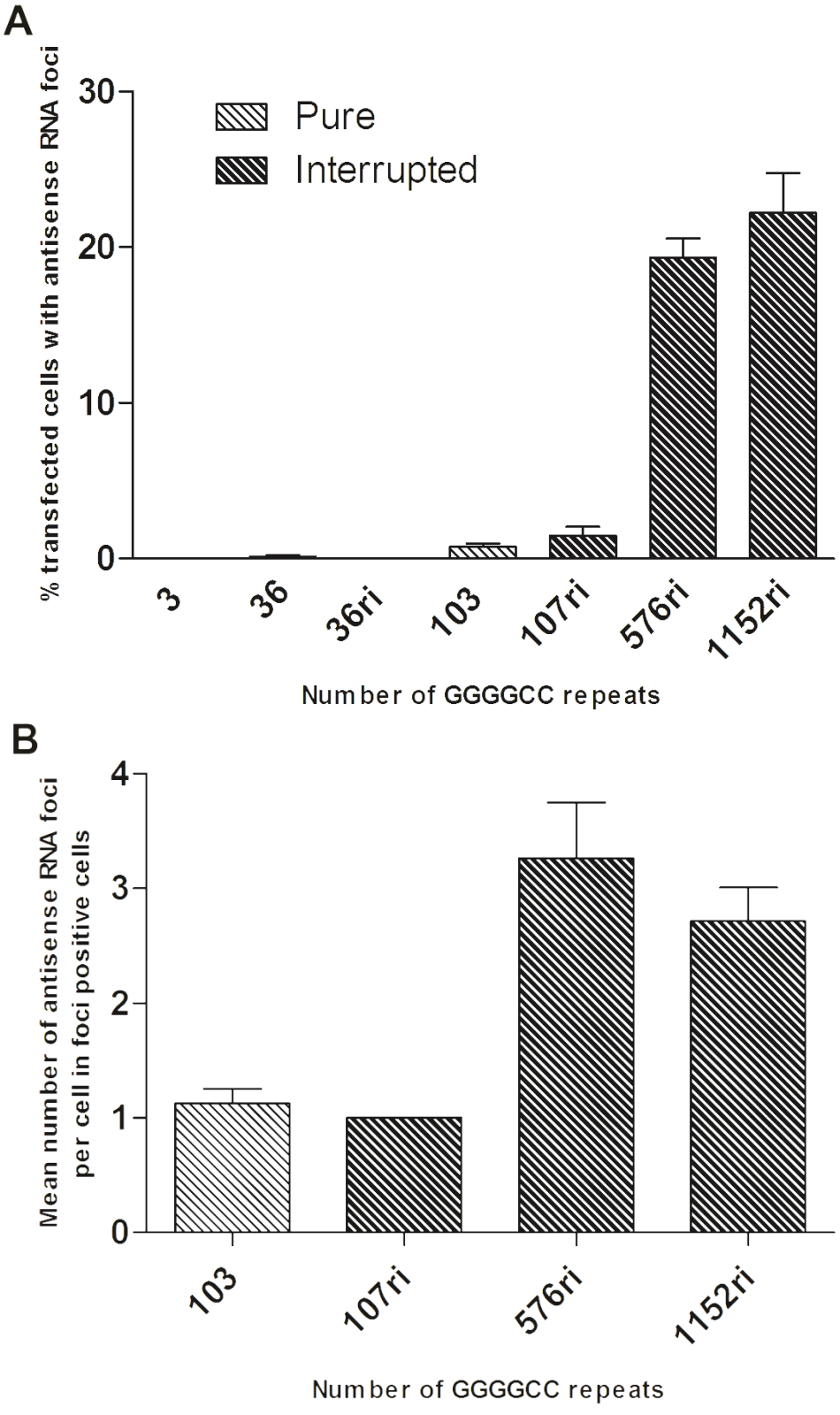


Figure 3.17 Quantification of antisense RNA foci formation in SH-SY5Y cells

A) The percentage of transfected cells that contain antisense RNA foci in SH-SY5Y cells transfected with a subset of different length pure (light bars) and interrupted (dark bars) repeat constructs B) Mean number of antisense RNA foci per cell exhibited by transfected cells that contain antisense RNA foci in SH-SY5Y cells transfected with a subset of different length pure and interrupted repeat constructs. Error bars represent standard error of mean (SEM). A minimum of 3 experimental replicates were carried out for each repeat size. r = pure repeats, ri = interrupted repeats.

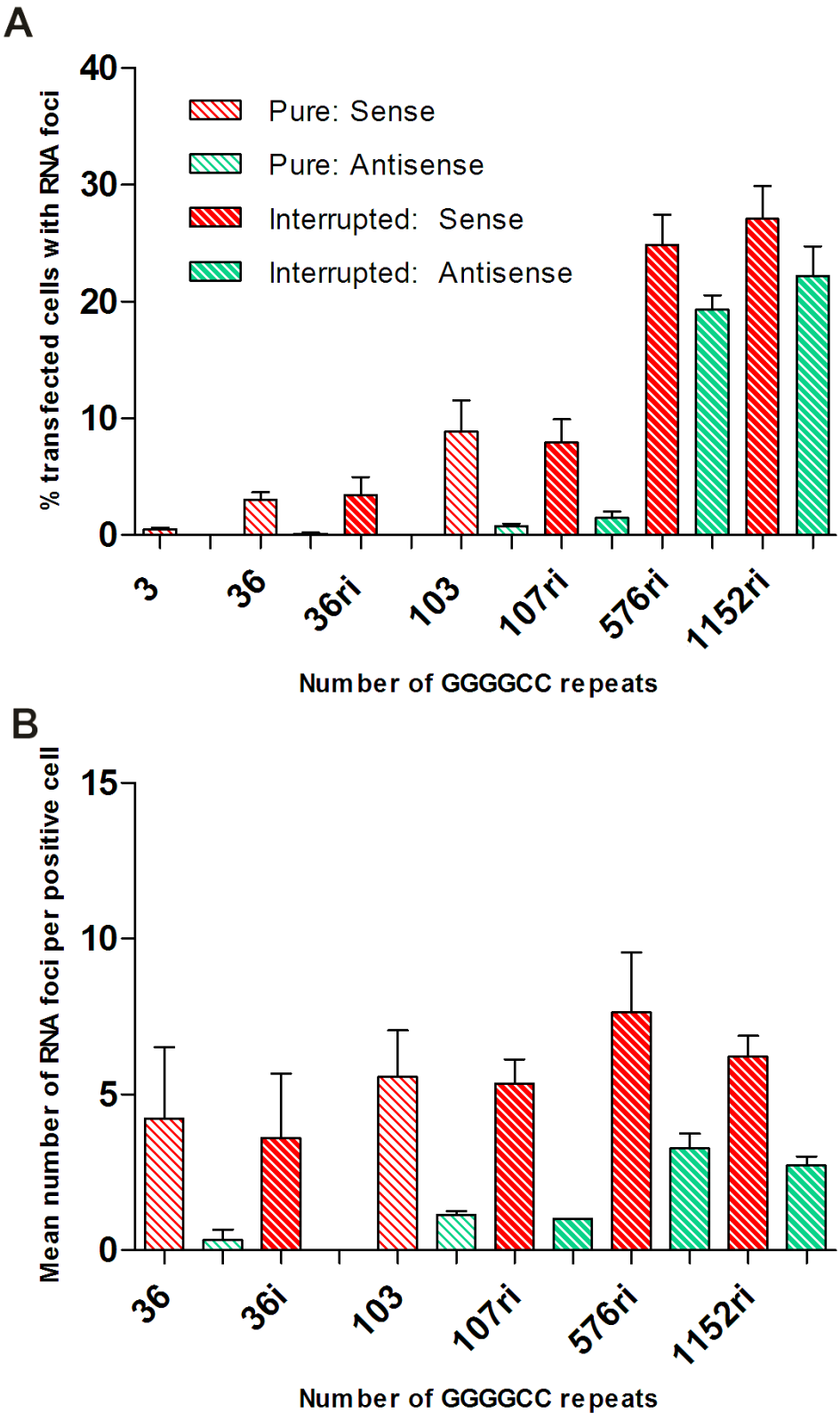


Figure 3.18 Comparison of sense and antisense RNA foci formation in SH-SY5Y cells

A) The percentage of transfected cells that contain sense (red bars) or antisense (green bars) RNA foci in SH-SY5Y cells transfected with a subset of different length pure (light bars) and interrupted (dark bars) repeat constructs B) Mean number of sense or antisense RNA foci per cell exhibited by transfected cells that contain RNA foci in SH-SY5Y cells transfected with a subset of different length pure and interrupted repeat constructs. Error bars represent standard error of mean (SEM). A minimum of 3 experimental replicates were carried out for each repeat size. r = pure repeats, ri = interrupted repeats.

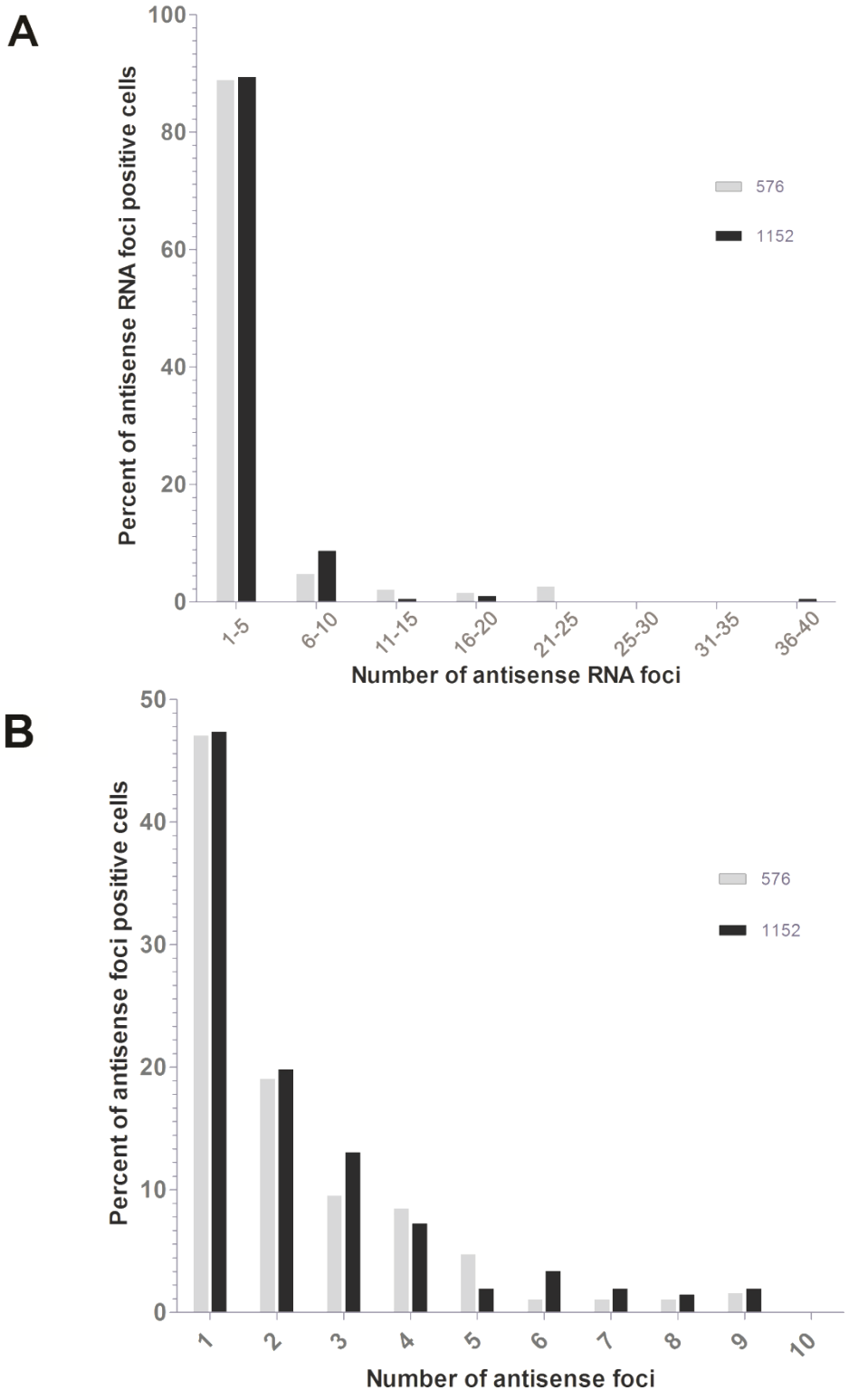


Figure 3.19 Frequency distribution of number of antisense RNA foci

Histogram displaying the proportion of cells exhibiting antisense RNA foci in SH-SY5Y cells transfected with different repeat constructs. . A) The number of antisense RNA foci expressed in all sense RNA foci positive cells was binned into groups of 5, with each bar representing the percentage of cells transfected with a given repeat construct that express a quantity of sense RNA foci within that bin range. B) The percentage of sense RNA foci positive cells exhibiting between 1 and 10 antisense RNA foci was examined in more detail.

numbers of foci were analysed, with around half of antisense RNA foci positive cells exhibiting one focus (Figure 3.19b). No difference in the distribution of antisense RNA foci number between different repeat lengths was apparent.

3.2.8 Expression of pure and RNA only repeats in a *Drosophila melanogaster* model

In collaboration with Prof Linda Partridge and researchers at the UCL Institute for Heathy Aging and the Max Planck Institute for Biology of Aging, a number of repeat expressing fly lines were generated. Generation of *Drosophila* lines was carried out by Dr. Sebastian Grönke and colleagues, while characterisation of the transgenic flies was carried out by Dr. Teresa Niccoli and colleagues. Pure and interrupted "RNA-only" constructs that had been subcloned into the pUASTattB vector were inserted into an identical genetic locus under the UAS promotor. Using the GMR–Gal4 driver, expression of the repeats was driven in the

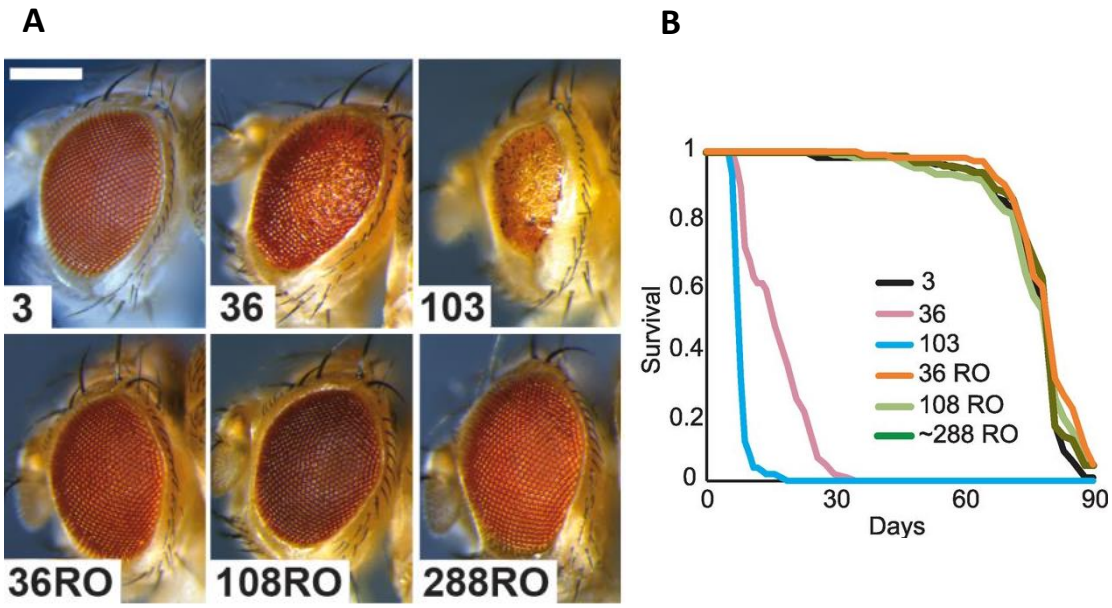


Figure 3.20 Expression of pure and RNA-only repeat constructs in *Drosophila*

A) Representative images of eyes from *Drosophila* incubated at 27 °C expressing pure 3r, 36r, 103r, and interrupted RNA-only (RO) 36ri, 108ri and 288ri repeats under the GMR-GAL4 driver. Scale bar represents 200 µm. B) Survival curves of flies expressing repeat constructs in adult neurons under the Elav-GeneSwitch driver.

fly eye. Examination of eye phenotype provides an easily assessable readout of degeneration. The expression of 36, 108 and 288 RNA-only repeats did not alter the configuration of the fly eye compared to a control line expressing 3 pure repeats (Figurre 3.20a). Conversely, both 36 and 103 pure repeats caused visible eye degeneration, with 103 pure repeats exhibiting a more severe degenerative phenotype than 36 pure repeats. These repeat constructs were also expressed in adult fly neurons using an inducible Elav-GeneSwitch driver. Flies expressing 36 or 103 pure repeats died within 30 days of induction, however no difference was seen between flies expressing 3 repeats and those expressing 36, 108 or 288 RNA-only repeats (Figure 3.20b). Once more, 103 pure repeats caused a more severe decrease in survival compared to 36 pure repeats. These data suggest that the primary species responsible for toxicity in these flies is DPR protein gain-of-function as opposed to RNA gain-of-function, and that the severity of toxicity rises with increased repeat length.

3.3 Discussion

3.3.1 Summary of results

In this chapter, we developed and characterised a new set of genetic tools that can be used to isolate RNA gain-of-function from DPR protein gain-of-function mechanisms of pathology caused by *C9orf72* hexanucleotide repeats.

A range of unique GGGGCC repeat constructs have been generated in this study. Pure GGGGCC repeats, which model both RNA and DPR protein gain-of-function mechanisms, were generated in sizes ranging from 3 to 103 repeats, while RNA-only interrupted GGGGCC repeats, which model only RNA gain-of-function mechanisms and preclude the production of DPR proteins, were generated in sizes ranging from 36 to 1152 repeats. Our largest RNA-only 1152 repeat construct falls within the range of repeat lengths observed in patients with the *C9orf72* repeat expansion. The array of different repeat sizes produced allows any phenomena observed in cells to be correlated with repeat length, facilitating the study of pathological phenotypes in a dose-dependent manner. A positive correlation was observed between repeat length and the percentage of transfected neuroblastoma cells containing both sense and antisense RNA foci, with no discernible difference in foci formation observed between pure and RNA-only constructs. However, no correlation was observed between repeat length and the number of RNA foci per cell exhibited by foci positive cells. Sense RNA foci were formed in a higher percentage of transfected cells than antisense RNA foci, with a higher number of sense foci formed in cells on average. Additionally, a higher repeat length threshold was observed for antisense RNA foci formation compared to sense RNA foci formation.

When these repeat constructs were expressed in a *Drosophila* model, pure repeats of 36 repeats and longer caused degeneration when expressed in the fly eye and adult neurons, while RNA-only repeats did not cause a discernible degenerative phenotype. More severe

degeneration was also observed in 103 pure repeats compared to 36 pure repeats, suggesting a correlation between repeat length and toxicity. The lack of toxicity observed in flies that produce GGGGCC repeat RNA but not DPR proteins suggests that RNA gain-of-function is not the primary toxic mechanism in this *Drosophila* model.

3.3.2 Cloning strategy for GGGGCC repeats

Generating these DNA repeat constructs necessitated overcoming a number of the inherent properties exhibited by long GGGGCC repeats that hamper the use of standard cloning protocols. First, repetitive constructs frequently undergo deletions when transformed into *E. coli* during the cloning process. This is in contrast to repeat instability in in human repeat disorders, in which a trend towards expansion is usually observed (McMurray, 2010). This difference may be due in part to the fact that in patients large repeat expansions are more often observed in non-dividing cells, in which the repair of single strand breaks is thought to lead to expansion, whereas in rapidly dividing cells such as *E. coli*, deletions more often occur due to errors such as template slippage and recombination during DNA replication (McMurray, 2010). Indeed, while deletions were frequently observed during the cloning of our repeat constructs, particularly in our pure repeat constructs and at higher repeat lengths, repeat expansions were very rarely observed during the cloning process.

To minimise the instability of our repeat constructs, we used a *recA* recombinase deficient *Stbl3 E.coli* strain. Transformed bacteria were also incubated at 30°C to maintain bacterial growth within log phase, as stationary growth phase has been observed to increase the frequency of repeat deletion (Parniewski et al., 2000). While these precautions allowed the construction of a maximum of 103 pure repeats, beyond this point constructs underwent large deletions, which we overcame with the use of interruptions to stabilise the repeat sequences. Interruptions have been shown to stabilise long repetitive sequences in a

number of myotonic dystrophy models without mitigating their pathological function. 480 CTG repeats with a 5 bp interruption every 20 repeats were observed to form RNA foci that sequestered MBLN1 homolog *muscleblind* protein in a *Drosophila* model, resulting in muscle wasting and degeneration in the fly eye (de Haro et al., 2006). In addition, 960 CTG repeats with a 5 bp interruption every 20 repeats was expressed in a mouse model, and recapitulated many aspects of myotonic dystrophy including myotonia, dysregulated alternative splicing, severe muscle wasting resulting in a functional impairment in treadmill tests, and co-localisation of MBLN1 to CUG RNA foci in skeletal muscle (Orengo et al., 2008). One study into the effect of interruptions on GAA repeats, present in Friedreich's ataxia, found that the inclusion of interruptions that comprised of more than 4 % of the repeat sequence substantially increased its stability: in this study a pure repeat sequence was reduced on average to 8 % of its original length after one re-cultivation in *E. coli*, while a sequence of a similar length with 4 % interruption content was only reduced to 85 % of its original size on average (Sakamoto et al., 2001). While increasing the interruption content was observed to further increase repeat stability over numerous re-cultivations, an interruption content of over 11 % disrupted the formation of secondary structure (Sakamoto et al., 2001). Our interrupted GGGGCC constructs were therefore designed with interruptions comprising 7.7 % of the total sequence, which was found to facilitate the construction of over 1000 repeats without disrupting RNA foci or G-quadruplex formation.

3.3.3 Investigating RNA foci formation in a SH-SY5Y cell culture model

There were a number of advantages to using a transiently transfected over-expression model in this study. Primarily, it provided a method for rapidly screening a wealth of different repeat constructs which facilitated a detailed examination of the effect of repeat dose on sense and antisense RNA foci formation. Furthermore, this cell model was a useful initial test to identify which constructs warranted further study in other models, such as in *Drosophila.* This model also served as an important proof of concept, demonstrating that

pure and interrupted constructs exhibit both sense and antisense foci formation to the same extent. This supports the notion that the interruptions do not disrupt the initiation of RNA gain-of-function mechanisms by these repeats, however further studies are needed to determine whether both pure and interrupted repeat constructs exhibit the same sequestration of RNA binding proteins.

Evidently, however, there are a number of ways in which the cells in this culture system differ from the cells in patients with *C9orf72* FTD/ALS, which should be taken into consideration when interpreting the results in this chapter. As *C9orf72* FTD/ALS principally affects the CNS, a neuronal-like cell line SH-SY5Y was chosen, which was originally derived from a human metastatic bone tumour biopsy (Biedler et al., 1973). This line expresses a number of immature neuronal markers and exhibits a neuroblastoma-like morphology, including the formation of short neurites (Kovalevich and Langford, 2013). In contrast to primary neuronal cultures, unlimited numbers of these immortalised cells can be grown in the lab with relative ease, with low culture to culture variability (Kovalevich and Langford, 2013). However, the post-mitotic nature of neurons is posited to be one of the reasons why the CNS is particularly susceptible to degeneration caused by genetic mutations which are ubiquitously expressed across cell types. In addition, the expression of repetitive DNA differs in a number of ways between patient neurons and our SH-SY5Y cell model. Firstly, repeat DNA is over-expressed under a strong CMV promotor in this cell model, with multiple vector copies potentially taken up in each transfected cell, which is likely to increase the burden of repetitive DNA in this cell culture model compared to patients. In addition, the time course for which the repeats are expressed in cells also differs greatly in this cell model from *in vivo*. While the *C9orf72* repeat expansion occurs in patients over an entire lifetime, in our cell system the repeats were expressed over a much shorter time period of 24 hours, meaning that we cannot study the long term effects that the repeats cause in patient tissues. Furthermore, *in vitro* culture is a stressful environment for cells,

which will differ greatly from a cell's *in vivo* niche, and will therefore alter the response of a cell to the repeats. All of these factors will affect the expression and behaviour of the repeat constructs to an extent, and it is therefore important to use cell culture models in conjunction with other studies, such as the *in vivo Drosophila* model we developed using the same constructs, as well as comparing observations made in cell models to those made in patient tissue where possible.

3.3.4 A comparison of RNA foci formation in cell culture, patient brain and iPSCs

Sense RNA foci formation was first described in one of the initial studies reporting the existence of a GGGGCC repeat in *C9orf72*, with 25 % of neurons in the frontal and motor cortices observed to exhibit foci (DeJesus-Hernandez et al., 2011). However, an ensuing study was unable to replicate this finding, bringing the existence of RNA foci in *C9orf72* patients into question (Simón-Sánchez et al., 2012). Subsequently however, our group as well as others have developed optimised higher sensitivity FISH protocols, which clearly indicate the presence of both sense and antisense RNA foci formation in *C9orf72* patient brain (Donnelly et al., 2013; Gendron et al., 2013; Lagier-Tourenne et al., 2013; Mizielinska et al., 2013; Zu et al., 2013). A sizeable degree of variation can be seen between the exact frequencies of sense and antisense RNA foci observed in different studies, which may be in part due to a high patient to patient variability as well as variations between the sensitivities of the FISH protocols used by different groups.

Despite the disparities between *C9orf72* patient neurons and the cell system explored in this chapter, there are some striking similarities in RNA foci formation observed in these two contexts. Studies examining *C9orf72* patient brain have found that sense and antisense RNA foci are not exhibited in every neuron, but in a subset of the population (Cooper-Knock et al., 2015b, 2014; Gendron et al., 2013; Lagier-Tourenne et al., 2013; Lee et al., 2013; Mizielinska et al., 2013; Zu et al., 2013), which is also observed in our cell

culture model. It is unclear at present which factors are responsible for promoting RNA foci in one cell, but not a neighbouring cell. In the frontal cortex, cerebellar granule cell layer, and hippocampus, a higher proportion of cells were observed to exhibit sense RNA foci in comparison to antisense foci (Gendron et al., 2013; Mizielinska et al., 2013; Zu et al., 2013), as was also observed in my SH-SY5Y cell model. As we would anticipate transcription of the *C9orf72* locus to be promoted predominantly in the sense direction, a higher frequency of sense RNA foci is perhaps to be expected. Conversely, however, one study in *C9orf72* ALS patients found that in the motor cortex and purkinje cell layer of the cerebellum, antisense RNA foci were more prevalent than sense foci (Cooper-Knock et al., 2015b). Additionally, antisense foci burden was observed to correlate with TDP-43 cytoplasmic localisation in *C9orf72* ALS motor neurons (Cooper-Knock et al., 2015b), while no correlation was observed between the presence of TDP-43 aggregates and the formation of sense or antisense RNA foci in the frontal cortex, cerebellum or hippocampus of *C9orf72* FTD patients (Mizielinska et al., 2013). These data suggest that different areas of the brain may be differentially affected by sense or antisense foci, which could potentially be a contributing factor to whether a patient presents with ALS or FTD pathology.

A number of different groups have examined sense and antisense RNA foci formation in the frontal cortex which is summarised in Table 3.3, with between 10-37 % of neurons forming sense RNA foci, and 8-26 % of neurons forming antisense RNA foci (DeJesus-Hernandez et al., 2011; Gendron et al., 2013; Mizielinska et al., 2013; Zu et al., 2013). A study from our group also noted a positive correlation between sense RNA foci burden in the frontal cortex and age of onset, consistent with a role for sense RNA foci in *C9orf72* FTD pathology (Mizielinska et al., 2013). The percentage of cells exhibiting RNA foci in our largest 1152ri construct, which most closely resembles the repeat length observed in patients with the expansion, falls towards the upper bound of the percentages of neurons observed to produce RNA foci by different studies in *C9orf72* FTD frontal cortex, with 27 ± 2.9 % of

Reference	Cohort	Sense foci	Antisense foci
Dejesus-Hernandez et al. (2012)	FTD	25 %	N/A
Gendron et al. (2014)	FTD/ALS	10 %	8 %
Mizielinska et al. (2014)	FTD	37 %	26 %
Zu et al. (2014)	ALS	15 %	10 %

Table 3.3 Percentage of *C9orf72* patient frontal cortex neurons detected to be positive for sense or antisense foci in different studies.

transfected cells exhibiting sense RNA foci, and 22 ± 2.5 % of cells exhibiting antisense RNA foci.

Induced pluripotent stem cells (iPSCs) and neurons differentiated from these cells (iPSNs) have also been observed to form sense RNA foci (Almeida et al., 2013; Donnelly et al., 2013; Sareen et al., 2013). As seen in both *C9orf72* FTD/ALS patient brain tissue and in our SH-SY5Y cell model, a large proportion of cells did not exhibit RNA foci in these studies. The percentages of cells observed to express sense RNA foci varied greatly both between studies, and also between iPSC lines within each study. One study reported sense RNA foci formation in between 20 % and 60 % of iPSCs depending on the line, including in lines derived from the same carrier, suggesting that numerous different environmental as well as genetic factors may interact to affect RNA foci formation (Almeida et al., 2013). Work in *C9orf72* ALS iPSC lines has additionally implicated that the interaction between GGGGCC repeat RNA and RNA binding proteins may be important in the formation of RNA foci, as knockdown of GGGGCC RNA binding protein ADARB2 in *C9orf72* ALS iPSCs resulted in a significant reduction in the number of sense foci positive cells (Donnelly et al., 2013).

When RNA foci number per cell was quantified in our SH-SY5Y cell model, the majority of RNA foci positive cells were observed to exhibit just one focus, with a smaller proportion exhibiting a few foci, and a minority of cells exhibiting a large number of foci. This

distribution profile has also been observed in *C9orf72* FTD patient frontal cortex (Mizielinska et al., 2013), as well as in *C9orf72* ALS patient fibroblasts and iPSC-derived motor neurons (Lagier-Tourenne et al., 2013; Sareen et al., 2013). However, while a higher average number of sense RNA foci were formed per foci-positive cell compared to antisense RNA foci in our SH-SY5Y cell model, the converse was observed in *C9orf72* FTD patient brain (Mizielinska et al., 2013), as well as in patient fibroblasts (Lagier-Tourenne et al., 2013). Individual neurons in *C9orf72* FTD patient brain tissue were observed to express up to 60 antisense RNA foci, with a maximum of 10 sense foci seen in any one cell (Mizielinska et al., 2013). Additionally, significantly more antisense foci than sense foci were detected on average per foci positive neuron (Mizielinska et al., 2013). Likewise, a maximum of 40 sense RNA foci were observed in *C9orf72* ALS patient fibroblasts, while up to 90 antisense foci in a single cell were observed (Lagier-Tourenne et al., 2013). In contrast, a maximum of 58 sense foci were detected in SY-SY5Y cells compared to a maximum 39 antisense foci, and 1152ri transfected cells exhibited an average of 6.2 sense foci per positive cell compared to 2.7 antisense foci. This may be due to the overexpression used in this model, in which transcription is driven much more strongly in the sense direction using a CMV promotor, in contrast to the repeats in patient cells which exist in a broader genomic context, thus potentially causing differences in transcription. It is also possible that degradation of antisense RNA foci is less efficient than sense RNA foci, which would lead to an accumulation of antisense RNA foci over time within patient cells.

3.3.5 GGGGCC repeat length and RNA foci formation

An array of different length repeat constructs were generated as a consequence of the RDL cloning strategy used to build long repeats. This enabled the correlation of repeat length against RNA foci formation in our SH-SY5Y cell model, and produced some interesting results. One finding was that while the percentage of transfected cells exhibiting sense or antisense foci was significantly correlated with repeat length, the average number of foci

exhibited by a positive cell remained the same. A similar percentage of the RNA-foci-positive cell population was observed to exhibit any given number of foci independent of the length of the repeat construct. This is exemplified by the observation that several cells transfected with 36 pure or interrupted repeats that exhibited over 30 foci. One interpretation of this result could be that increased repeat length promotes the initial nucleation of RNA foci, but that once RNA aggregation has been initiated, other factors determine the extent of foci formation.

The relationship between repeat length and foci formation has been examined in several studies in patient-derived cells. One study examining RNA foci formation in patient-derived fibroblast cultures found that while the percentage of cells exhibiting sense RNA foci in different lines varied between 15 and 45 %, there was no clear relationship evident between repeat length, patient age, or clinical diagnosis versus the percentage of cells that exhibited sense RNA foci (Lagier-Tourenne et al., 2013). Additionally, a study examining patient derived iPSCs and neurons differentiated from these cells did not observe a correlation between cell repeat length and either the percentage of sense RNA foci positive cells or the number of sense foci per cell (Almeida et al., 2013). This suggests that factors other than repeat length are likely to affect the extent of RNA foci formation in cells, however precise measurement of long GGGGCC repeats presents a technical challenge, and inaccuracies in repeat sizing could potentially mask a correlation between repeat length and cell phenotype. In a case homozygous for the repeat expansion (Fratta et al., 2013), both sense and antisense RNA foci were found to be more frequent in the frontal cortex than heterozygous cases, with more sense and antisense RNA foci also detected per cell on average (Mizielinska et al., 2013), arguing in favour of some dose dependent effect of the repeats on RNA foci formation in cells.

3.3.6 GGGGCC repeat length and patient disease course

The relationship between *C9orf72* repeat length and disease progression in patients is currently unclear. In a number of other repeat disorders, a very clear correlation is observed between repeat length and age of onset. The length of the CAG repeat expansion in Huntington's disease, for example, can be used to predict a statistical probability of disease onset at a given age (Langbehn et al., 2004). In myotonic dystrophy patients, inherited CTG repeat length was found to be predictive of disease onset, however, the high somatic instability of CTG repeats necessitates that inherited repeat length is estimated from the spread of repeat sizes present in patient cells (Morales et al., 2012). Additionally, the degree of somatic instability in myotonic dystrophy patients was also found to correlate inversely with disease onset, with more rapid expansion of the repeat leading to an earlier disease onset (Morales et al., 2012).

C9orf72 hexanucleotide repeats have been observed to be highly unstable. In one report, a patient with around 90 GGGGCC repeats in the blood was found to have an expansion of thousands of repeats when tested in the CNS, indicating a high potential for differences in repeat sizes between different tissues (Fratta et al., 2015). A study into the correlation between *C9orf72* repeat length and disease phenotype by van Blitterswijk et al. (2013) also found a high degree of somatic instability in patient tissue, with differences in repeat size observed in different brain areas as well as in the blood. Counterintuitively, this report found that repeat length correlates positively with age of onset in the frontal cortex of FTD patients, with shorter repeat lengths associated with an earlier disease onset (van Blitterswijk et al., 2013). This correlation was not observed in other brain regions or in other disease subtypes, and the implications underlying this trend are presently unclear. This difference in repeat length could be hypothesised to be caused by the somatic instability of the repeat sequence and the propensity for repeats to expand over time, with longer repeats therefore observed in older patients, however another study found no

significant correlation between *C9orf72* expansion size in the cerebellum and age at autopsy (Suh et al., 2015). Additionally, repeat length was observed to be significantly lower in cerebellum compared to those in the frontal cortex and blood, and patients with cerebellar repeat lengths in the top 25th percentile were observed to have significantly reduced disease duration compared to the total cohort (van Blitterswijk et al., 2013). While the cerebellum is not a major site of pathology in FTD or ALS, it is possible that the smaller average repeat length reflects a lower degree of somatic instability in this area and therefore approximates inherited repeat length which was seen to be predictive of disease onset in myotonic dystrophy (Morales et al., 2012). Several studies have also looked into the relationship between *C9orf72* repeat length and disease subtype. While one report observed no significant difference between repeat length in the frontal cortex, cerebellum or blood between patients with MND, FTD or FTD/MND (van Blitterswijk et al., 2013), work by another group found that *C9orf72* repeat length in the blood was shorter in FTD patients than ALS patients (Suh et al., 2015), suggesting that repeat length could potentially a determining factor for clinical disease presentation in patients. Furthermore, *C9orf72* expansion size was observed to be significantly associated with diminished disease duration for patients with FTD but not ALS (Suh et al., 2015), raising the question of whether repeat length differentially affects different disease subtypes. However, the relationship between GGGGCC repeat length and disease progression remains unclear and further studies are needed to fully comprehend these conflicting reports.

There must exist some form of relationship between repeat length and disease onset, as we know that healthy individuals most often exhibit 2 GGGGCC repeats, with individuals exhibiting no signs of pathology reported up to at least 32 repeats (Beck et al., 2013; Ratti et al., 2012). Some studies have reported the presence of psychiatric symptoms and cognitive deterioration in individuals with intermediate repeat lengths between 20 and 30 units (Byrne et al., 2014; Gómez-Tortosa et al., 2013), however others have reported that

while these intermediate repeat lengths were observed to affect transcription at the *C9orf72* locus, no association was observed between these lengths and disease (van der Zee et al., 2013). These different observations could be explained by differing repeat lengths between different tissues due to somatic repeat instability. One case report also described a patient with approximately 70 repeats in blood, who did not exhibit signs of degeneration or RNA foci in fibroblasts at age 89, while offspring from this patient were observed to carry an expansion of approximately 1750 repeats and exhibited RNA foci (Xi et al., 2015a). The relationship of repeat length and disease onset is further complicated by the incomplete penetrance of the disease, with individuals carrying large expansions occasionally living neurologically healthy lives into old age, and disease onset observed in some patients in their 80s (Majounie et al., 2012b).

A number of factors may underlie the ambiguous correlation with of disease phenotype with *C9orf72* repeat length. Firstly, there is an absence of patients with repeat lengths in the low hundreds, meaning that any differences in disease phenotype caused by a difference in repeat length within this range cannot be observed. It is also likely that a number of other genetic and environmental factors interact to affect disease progression, which could mask the contribution of repeat length and render it an inadequate predictor of disease prognosis. Another possibility is that repeat length exists as a binary switch, with the existence of a discrete repeat length threshold beyond which the full pathology occurs. This is an interesting possibility as it might also suggest the existence of a binary switch within the molecular disease mechanism, in which a tipping point is reached in the cell leading to full pathogenesis. However, methods for more accurately determining repeat length in patient brain may yet reveal clearer correlations between repeat length and patient disease severity in patients of different disease subtypes that cannot be detected at present.

3.3.7 Investigating RNA gain-of-function toxicity in a *Drosophila* model

The constructs generated in this work provide a useful tool that can be applied to a range of different models. In a collaboration between Linda Partridge's group and our own, a range of pure and RNA-only were expressed in *Drosophila melanogaster*, in the eye and in adult neurons (Mizielinska et al., 2014). Sense repeat RNA transcripts were detected in flies expressing pure and RNA-only constructs. Additionally, sense RNA foci were observed in the nuclei of the *Drosophila* salivary glands, which were examined due to their large nuclei which facilitated foci visualisation. However, no expression of antisense transcripts was detected in flies, meaning that the contribution of antisense RNA gain-of-function was not examined in this model. There is the potential for the effect of antisense RNA to be determined in future studies using promotor driven GGCCCC antisense repeat constructs.

RAN translation was observed to be successfully attenuated by the RNA-only constructs when expressed in the *Drosophila* model. Poly(GR) expression was assessed using dot-blot analysis, with 36 and 103 pure repeats found to undergo translation to produce poly(GR) protein, but not 36, 108 or 288 repeat RNA-only constructs. Additionally, western blot analysis revealed the presence of poly(GP) in 36 and 103 pure repeats, but not in 36, 108 or 288 repeat RNA-only constructs. This suggests that our construct designs are effective in inhibiting RAN translation in the RNA-only repeat constructs used in this model, while allowing the induction of RAN translation in the pure repeats. Due to the design constraints of the interrupted RNA-only constructs, which limited the percentage of the sequence that could comprise of interruptions, one stop codon occurs every 36 repeats in any one frame in these constructs. RAN translation has been found to be length dependent; it is not yet certain what the minimum number of repeats that can cause RAN translation is, but a low level of RAN translation has been observed in constructs from 38 repeats in length when expressed in HEK293 cells (Mori et al., 2013c). However, the

absence of DPR proteins detected from *Drosophila* expressing RNA-only constructs suggests that our design is effective in inhibiting RAN translation in this model.

36 and 103 pure repeats were observed to cause degeneration when expressed in the *Drosophila* eye as well as a reduction in survival when expressed in adult neurons. However, no such toxicity was observed in flies expressing 36, 108 or 288 RNA-only repeats. This suggests that, at least in this *Drosophila* model, sense RNA gain-of-function is not the predominant mechanism of toxicity as constructs that produce both repetitive RNA and DPR proteins result in degeneration, but constructs that produce only repetitive RNA and no DPR proteins do not. This additionally implicates DPR proteins in the observed degeneration, however, it does not preclude a role for RNA toxicity. Thus far, the greater instability of the larger 576 and 1152 RNA-only constructs has precluded the production of flies with longer repeat lengths. However, it would be interesting to investigate whether any RNA gain-of-function toxicity becomes apparent at longer repeat lengths.

Chapter 4: Investigating dipeptide repeat protein gain-of-function mechanisms in *C9orf72* FTD/ALS

4.1 Introduction

In the previous results chapter, the RNA gain-of-function mechanisms of pathology caused by the *C9orf72* repeat expansion were assessed in cell culture and *Drosophila melanogaster* model systems. While our pure GGGGCC repeat constructs, which produce both RNA and DPR proteins, caused acute toxicity when expressed in *Drosophila*, the expression of RNA-only constructs did not cause observable degeneration in this model system, thus implicating a role for DPR proteins in mediating GGGGCC repeat expansion toxicity. We therefore sought to characterise the role DPR proteins play in *C9orf72* FTD/ALS.

In this chapter, we detail the design and construction of new 'protein-only' DNA constructs that code individually for each of the five DPR proteins in a number of different repeat lengths in order to determine their pathological role. These protein-only constructs together with our pure and RNA-only DNA constructs form a powerful molecular toolkit, enabling the comprehensive dissection of the different gain-of-function mechanisms at work in *C9orf72* FTD/ALS in a repeat length dependent manner.

Protein-only constructs were transfected into a HEK293T cell line, and the subcellular localisation of GFP tagged and untagged DPR proteins of different lengths was characterised with immunofluorescent staining. A number of these protein-only constructs were then subcloned into a *Drosophila* vector and expressed in a fly model as part of our collaboration with Prof. Linda Partridge at the UCL Institute for Healthy Ageing and Max Planck Institute for Biology of Aging, in order to study the effect of individual DPR proteins in this system.

The role of DPR proteins in human brain has recently been subject to a great deal of debate. While a correlation has been reported between DPR protein inclusion burden and age of disease onset in *C9orf72* FTD/ALS patients, a lack of correlation has also been observed between the degree of degeneration in different brain regions and the level of DPR protein burden in each area (Davidson et al., 2014; Gomez-Deza et al., 2015; Mackenzie et al., 2013; Schludi et al., 2015). Some studies have reported a distinct DPR protein burden profile in patients with different disease subtypes (Gendron et al., 2015; Schludi et al., 2015), but other studies have failed to replicate this (Mackenzie et al., 2015).

To help clarify the relevance of DPR protein gain-of-function pathology to human *C9orf72* patients in greater detail, we designed a fully quantitative, high throughput and unbiased method for determining DPR protein aggregate frequency as detected by immunohistochemistry in human brain sections, in the form of an image analysis protocol. We then tested this protocol by calculating the percentage of DPR protein inclusion positive cells in *C9orf72* FTD post-mortem cerebellum.

4.2 Results

4.2.1 Cloning protein only constructs

To further dissect different gain-of-function mechanisms in *C9orf72* FTD/ALS, we generated 'protein-only' DNA constructs that yield each individual DPR protein in isolation, and do not produce GGGGCC RNA. DNA sequences were designed that use alternative codons to those found in GGGGCC repeats to code for each DPR protein separately. The protein-only DNA sequences were non-repetitive and had a lower GC content than pure GGGGCC repeats, allowing for the synthesis of different length repeat constructs without the use of the RDL cloning strategy. GFP tagged $(DPR)_6$ constructs were made by ligating oligonucleotides coding for six repeats of each of the DPR proteins (Table 4.1) into a pAcGFP1-C1 mammalian expression vector in frame with GFP. Sequences encoding 36 and 100 dipeptide repeats were synthesised by GeneArt in a pMK-RQ vector, and were subcloned into pcDNA3.1+ mammalian expression vector and pUASTaattB *Drosophila* expression vector to yield untagged $(DPR)_{36}$ and $(DPR)_{100}$ constructs. $(DPR)_6$-GFP and $(DPR)_{36}$ constructs were generated for all five DPR proteins, and $(DPR)_{100}$ constructs were generated for GA, GR, PR and AP, however a stable $(GP)_{100}$ construct could not be generated with the

Dipeptide	Sequence	Amino acid	GC-only codons	Alternative Codons
$(Gly\text{-}Ala)_6$	GGTGCAGGAGCTGGTGCAGGAGCTGGTGCAGGAGCT	Glycine	GGC GGG	GGT GGA
$(Gly\text{-}Pro)_6$	GGTCCAGGACCTGGTCCAGGACCTGGTCCAGGACCT	Alanine	GCC GCG	GCT GCA
$(Gly\text{-}Arg)_6$	GGTAGAGGAAGAGGTAGAGGAAGAGGTAGAGGAAGA	Proline	CCC CCG	CCT CCA
$(Pro\text{-}Arg)_6$	CCTAGACCAAGACCTAGACCAAGACCTAGACCAAGA	Arginine	CGC CGG	CGT CGA AGA AGG
$(Ala\text{-}Pro)_6$	GCTCCAGCACCTGCTCCAGCACCTGCTCCAGCACCT			

Table 4.1 Sequence of $(DPR)_6$ repeats and available amino acid codons

alternative codons available. The resulting vectors were sequenced to check for mutations. No sequence changes were detected in the generated constructs with the exception of $(AP)_6$.GFP, which underwent a modest expansion to $(AP)_7$-GFP.

4.2.2 Characterisation of DPR protein production in cell model

We next sought to characterise the production and subcellular localisation of protein produced by our different length DPR protein-only constructs. $(DPR)_{6/7}$-GFP, $(DPR)_{36}$ and $(DPR)_{100}$ constructs and empty pcDNA3.1+ vector were transiently transfected into the immortalised human embryonic kidney cell line HEK293T, which have been widely used in the literature to study DPR proteins and RAN translation in *C9orf72* FTD/ALS (May et al., 2014; Mori et al., 2013c; Tao et al., 2015; Yamakawa et al., 2014; Zu et al., 2013). $(DPR)_{36}$, $(DPR)_{100}$ and empty vector transfected cells were co-transfected with a ZsGreen co-transfection marker to identify transfected cells, while GFP tagged $(DPR)_{6/7}$-GFP constructs were transfected alone. Transfected cells underwent immunofluorescent staining with antibodies against the relevant DPR protein. An array of different antibodies reactive against each DPR protein were tested in a range of concentrations on cells transfected with each of the protein-only constructs as well as empty vector control. Antibodies and concentrations that yielded the most robust immunofluorescent signal in protein-only transfected cells compared to background staining in empty vector controls were selected to examine in more detail.

The co-localisation of antibody signal with GFP tag signal in $(DPR)_{6/7}$-GFP transfected cells was studied to investigate antibody specificity (Figures 4.1 to 4.5; Table 4.3). Antibodies against poly(GP), poly(GR), poly(AP) and poly(PR) identified diffuse cytoplasmic localisation of $(DPR)_6$ protein in GFP positive cells, with no discernible signal observed in empty vector transfected cells, indicating that these antibodies are specifically reactive to their

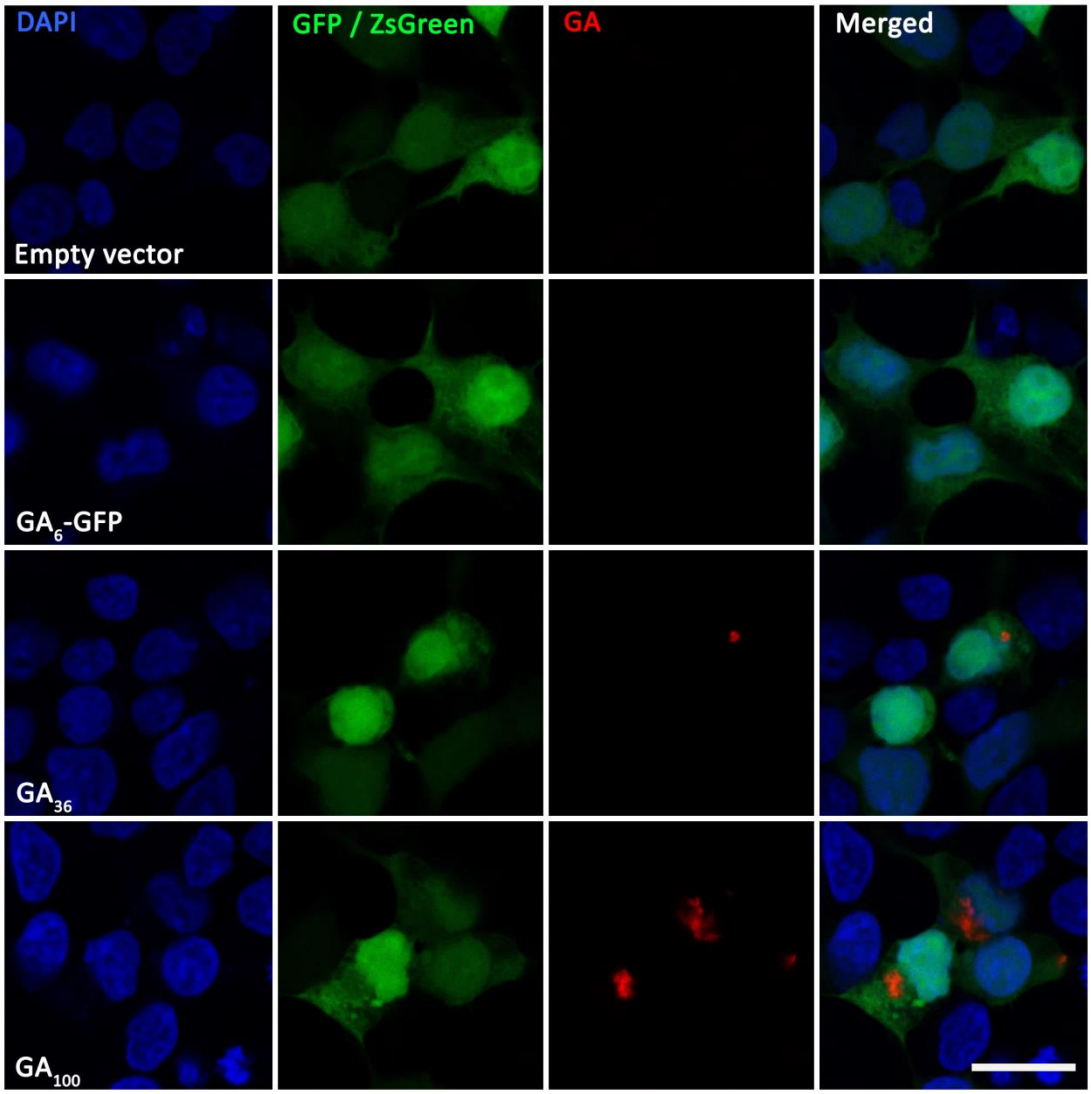


Figure 4.1 Subcellular localisation of protein produced by transiently expressed glycine-alanine constructs in HEK293T cells

Immunofluorescent images of HEK293T cells transfected with glycine-alanine (GA) protein-only constructs. Empty vector, GA_{36} and GA_{100} transfected cells were co-transfected with ZsGreen, shown in green. Green fluorescence in $(GA)_6$-GFP transfected cells represents GFP tagged construct. GA immunostain shown in red, DAPI nuclear stain shown in blue. Scale bar represents 20 µm. Images are representative of n = 3.

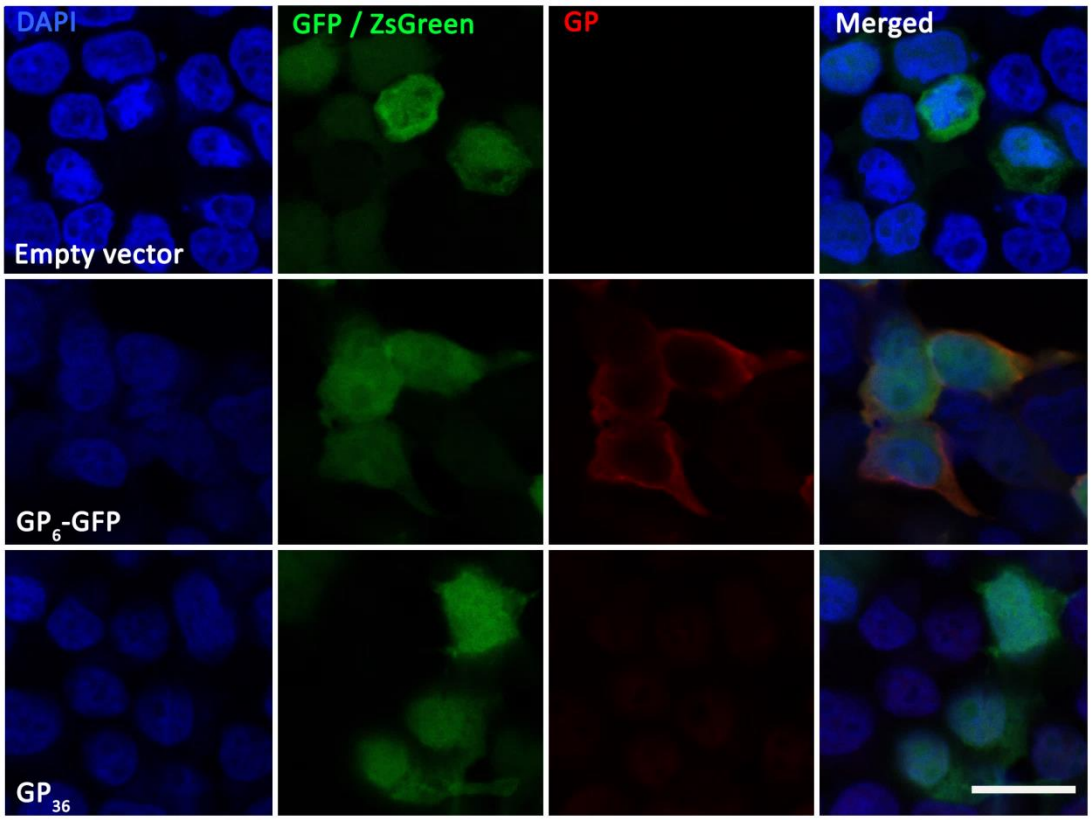


Figure 4.2 Subcellular localisation of protein produced by transiently expressed glycine-proline constructs in HEK293T cells

Immunofluorescent images of HEK293T cells transfected with glycine-proline (GP) protein-only constructs. Empty vector, GP_{36} and GP_{100} transfected cells were co-transfected with ZsGreen, shown in green. Green fluorescence in $(GP)_6$-GFP transfected cells represents GFP tagged construct. GP immunostain shown in red, DAPI nuclear stain shown in blue. Scale bar represents 20 µm. Images are representative of n = 3.

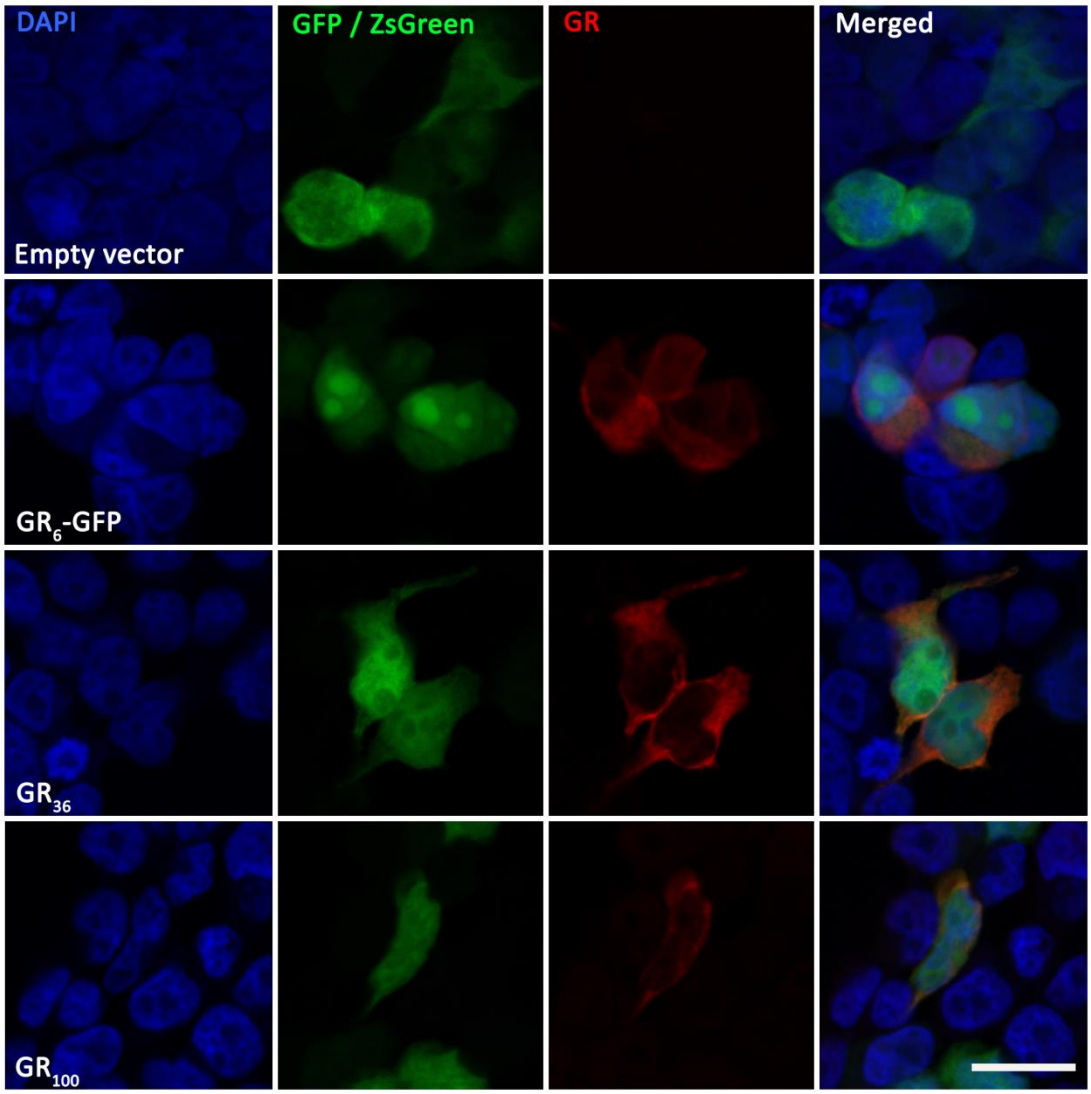


Figure 4.3 Subcellular localisation of protein produced by transiently expressed glycine-arginine constructs in HEK293T cells

Immunofluorescent images of HEK293T cells transfected with glycine-arginine (GR) protein-only constructs. Empty vector, GR_{36} and GR_{100} transfected cells were co-transfected with ZsGreen, shown in green. Green fluorescence in $(GR)_6$-GFP transfected cells represents GFP tagged construct. GR immunostain shown in red, DAPI nuclear stain shown in blue. Scale bar represents 20 µm. Images are representative of n = 3.

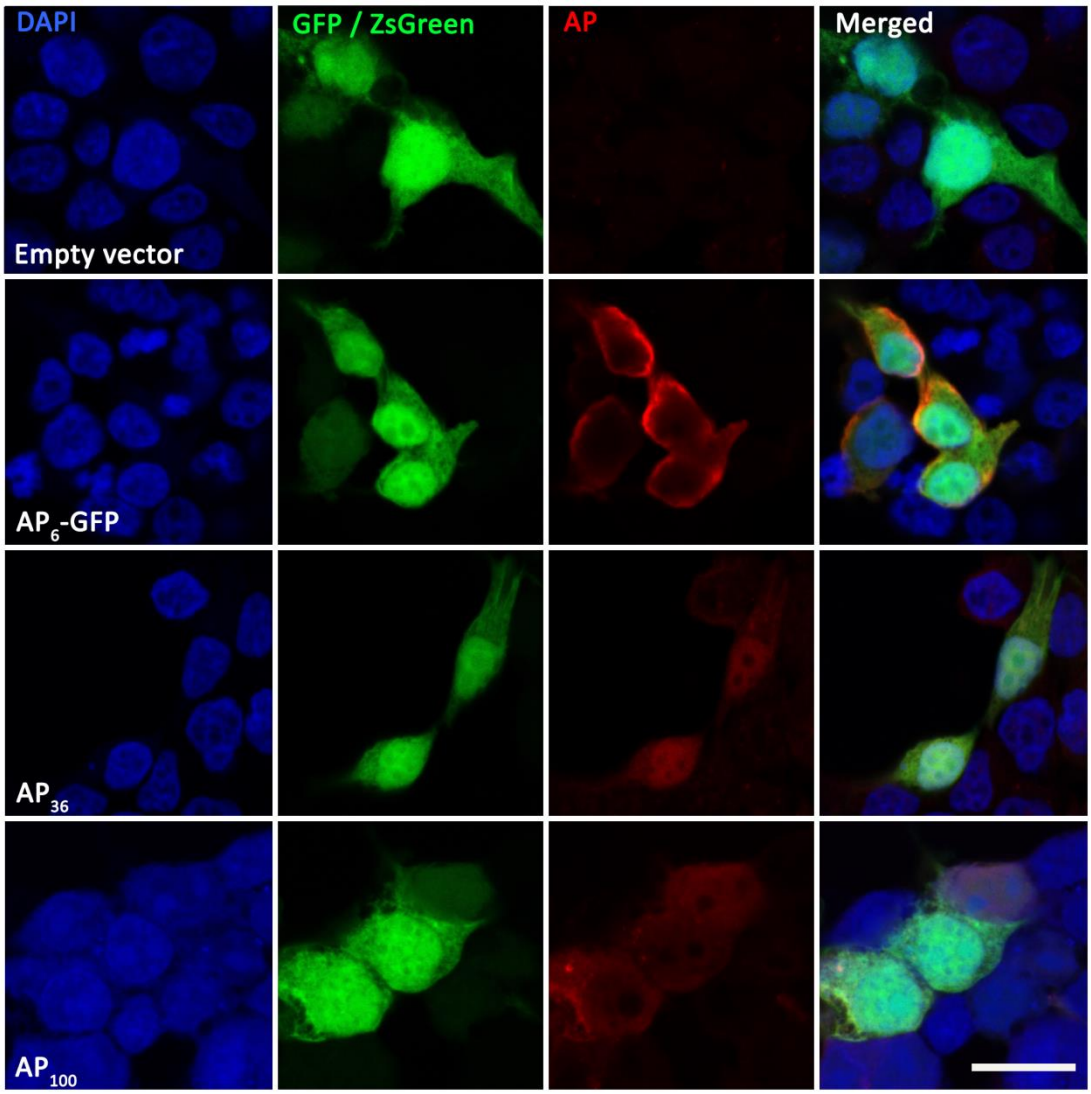


Figure 4.4 Subcellular localisation of protein produced by transiently expressed alanine-proline constructs in HEK293T cells

Immunofluorescent images of HEK293T cells transfected with alanine-proline (AP) protein-only constructs. Empty vector, AP_{36} and AP_{100} transfected cells were co-transfected with ZsGreen, shown in green. Green fluorescence in $(AP)_7$-GFP transfected cells represents GFP tagged construct. AP immunostain shown in red, DAPI nuclear stain shown in blue. Scale bar represents 20 µm. Images are representative of n = 3.

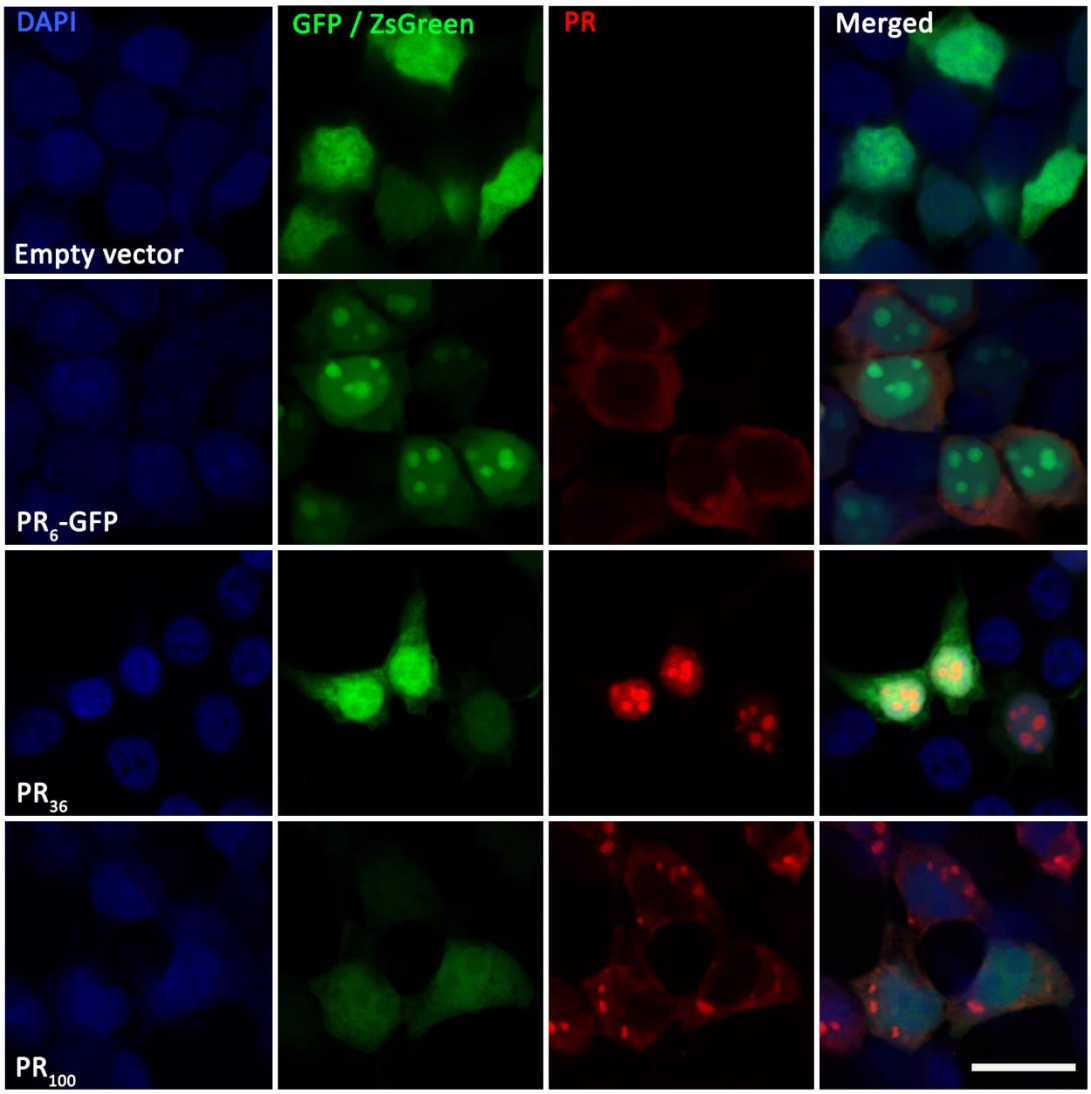


Figure 4.5 Subcellular localisation of transiently expressed proline-arginine constructs in HEK293T cells

Immunofluorescent images of HEK293T cells transfected with proline-arginine (PR) protein-only constructs. Empty vector, PR_{36} and PR_{100} transfected cells were co-transfected with ZsGreen, shown in green. Green fluorescence in $(PR)_6$-GFP transfected cells represents GFP tagged construct. PR immunostain shown in red, DAPI nuclear stain shown in blue. Scale bar represents 20 µm. Images are representative of n = 3.

Construct	Signal	GA		GR		GP		PR		AP	
		Nucleus	Cytoplasm	Nucleus	Cytoplasm	Nucleus	Cytoplasm	Nucleus	Cytoplasm	Nucleus	Cytoplasm
DPR_6GFP	GFP	+ Diffuse	+ Diffuse	+ Aggregates	+ Diffuse	+ Diffuse	+ Diffuse	+ Aggregates	+ Diffuse	+ Diffuse	+ Diffuse
	α-DPR	-	-	-	+ Diffuse	-	+ Diffuse	-	+ Diffuse	-	+ Diffuse
DPR_{36}	α-DPR	-	+ Small infrequent aggregates	-	+ Diffuse	-	-	+ Frequent aggregates	+ Infrequent aggregates	+ Diffuse	+ Diffuse
DPR_{100}	α-DPR	-	+ Large frequent aggregates	-	+ Diffuse	N/A	N/A	+ Infrequent aggregates	+ Frequent aggregates	+ Diffuse	+ Diffuse

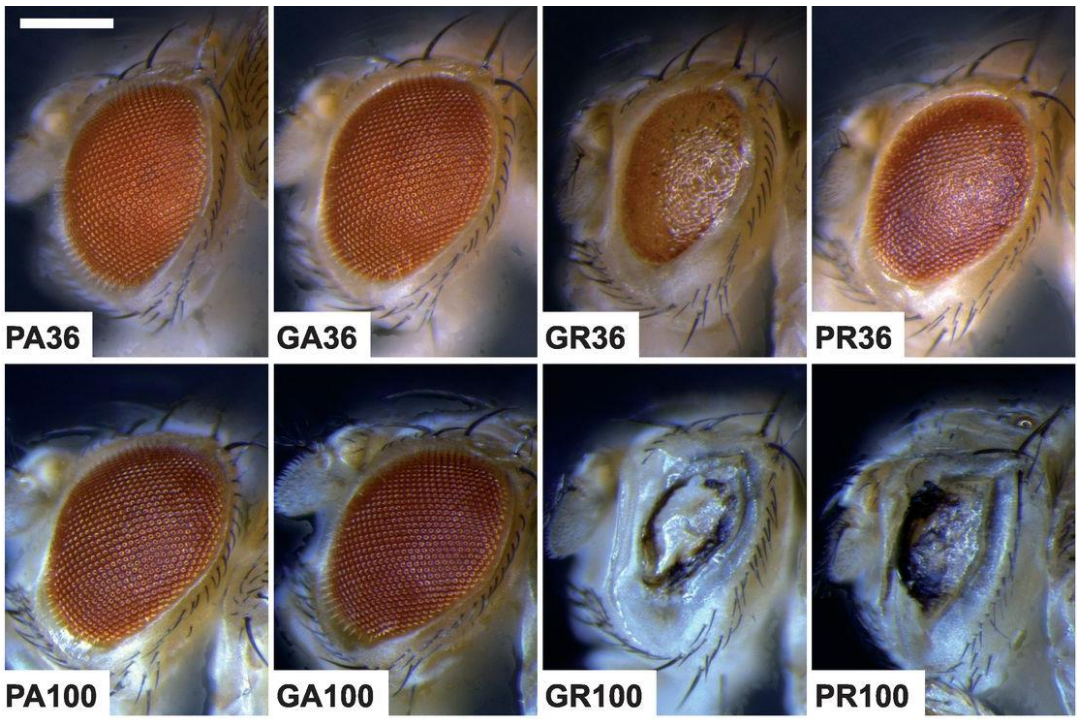
Table 4.2 Summary of subcellular DPR protein localisation in HEK293T cells transfected with protein-only constructs

respective DPR protein targets. However, while GFP was also observed to be present in the nucleus of $(DPR)_{6/7}$-GFP transfected cells, poly(GP), poly(GR), poly(AP) and poly(PR) were not detected in the nucleus of these cells by their respective antibodies.

While GFP signal was observed in $(GA)_6$-GFP transfected cells, no anti-poly(GA) antibody signal was visible in these cells (Figure 4.1). However, very frequent, large, star-like inclusions of poly(GA) were detected in $(GA)_{100}$ transfected cells, with smaller and very infrequent poly(GA) inclusions detected in $(GA)_{36}$ transfected cells, suggesting that this antibody recognises insoluble aggregates of poly(GA), and that longer poly(GA) repeat length promotes more extensive inclusion formation. Despite a lack of intranuclear anti-poly(GR) or anti-poly(PR) antibody signal, arginine containing $(GR)_6$-GFP and $(PR)_6$-GFP transfected cells exhibited GFP positive aggregates within the nucleus, which resembled the nucleolar localisation observed by other groups (Figure 4.3 and Figure 4.5) (Schludi et al., 2015; Tao et al., 2015; Yamakawa et al., 2014). $(PR)_{36}$ transfected cells were observed to form anti-poly(PR)-positive aggregates that were present primarily within the nucleus, however in $(PR)_{100}$ transfected cells, anti-poly(PR)-positive aggregates were mainly present in the cytoplasm, suggesting a differential cellular localisation of poly(PR), dependent on protein length. Conversely, only diffuse cytoplasmic anti-poly(GR) antibody staining was observed in $(GR)_{36}$ and $(GR)_{100}$ transfected cells. Cells transfected with $(AP)_{36}$ and $(AP)_{100}$ exhibited diffuse nuclear and cytoplasmic anti-poly(AP) antibody signal, despite a lack of nuclear anti-poly(AP) antibody signal observed in $(AP)_7$-GFP transfected cells (Figure 4.4). Finally, no anti-poly(GP) antibody signal was detectable in $(GP)_{36}$ transfected cells, despite a cytoplasmic anti-poly(GP) antibody signal detectable in $(GP)_6$-GFP transfected cells (Figure 4.2).

4.2.3 Effect of protein only constructs in a *Drosophila melanogaster* model

To investigate the effect of individual DPR proteins in a living model system, we again collaborated with Prof. Linda Partidge and researchers at the UCL Institute for Heathy Aging and the Max Planck Institute for Biology of Aging to generate repeat construct expressing *Drosophila melanogaster* lines (see section 3.2.8 for previous *Drosophila* collaboration work). Generation of *Drosophila* lines was carried out by Dr. Sebastian Grönke and colleagues, while characterisation of the transgenic flies was carried out by Dr. Teresa Niccoli and colleagues. $(DPR)_{36}$ and $(DPR)_{100}$ protein-only constructs that had been



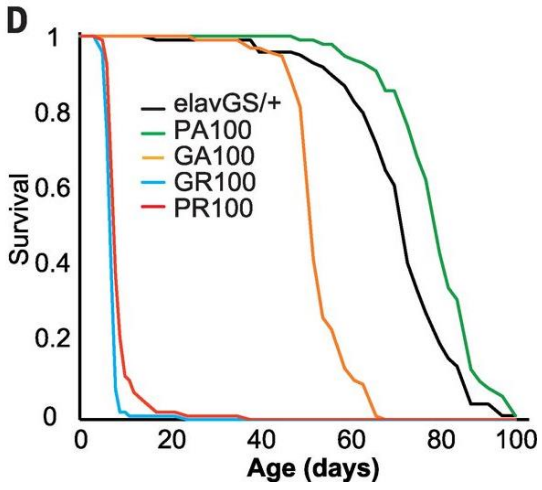


Figure 4.6 Expression of protein-only repeat constructs in *Drosophila*

A) Representative images of eyes from *Drosophila* expressing 36 or 100 repeats of proline-alanine (PA), glycine-alanine (GA), glycine-arginine (GR), and proline-arginine (PR) protein-only constructs under the GMR-GAL4 driver. Scale bar represents 200 µm. B) Survival curves of flies expressing 100 repeat protein-only constructs in adult neurons under the Elav-GeneSwitch driver, or expressing the driver alone (elavGS/+).

subcloned into pUASTattb were inserted into an identical genetic locus under the UAS promotor. Using the GMR–Gal4 driver, expression of the repeats was driven in the fly eye, which can be easily assessed for signs of degeneration. While no disruption to the morphology of the fly eye was apparent in flies expressing poly(AP) or poly(GA) protein-only constructs of 36 or 100 repeats in length, marked ocular degeneration was observed in flies expressing arginine containing poly(GR) or poly(PR) protein-only repeats, with severe deterioration of the ordered compound eye structure (Figure 4.6a). Expression of 100 GR or PR protein-only repeats resulted in a more severe phenotype than 36 protein-only repeats indicating that the toxicity of these repeats increases with repeat length. Protein-only repeats were additionally expressed in fly adult neurons using an inducible Elav-GeneSwitch driver. Flies expressing 100 PR or GR repeats exhibited an extreme reduction in survival, with lethality observed in the majority of flies expressing these constructs within 10 days of expression induction (Figure 4.6b). A more modest reduction in survival was also observed in flies expressing 100 GA protein-only repeats, and no loss of viability was observed in flies expressing 100 PA protein-only repeats. These data show that the arginine-containing DPR proteins poly(GR) and poly(PR) are sufficient to cause severe, repeat length dependent toxicity in this *Drosophila* model system, in contrast to the lack of toxicity observed from GGGGCC RNA alone (section 3.2.8).

4.2.4 Development of an image analysis protocol for DPR aggregate quantification in *C9orf72* FTD/ALS patient post-mortem brain

To determine the relevance of the DPR protein toxicity observed in our *Drosophila* models to *C9orf72* pathology in human patients, we investigated the distribution of DPR protein aggregates in post-mortem human *C9orf72* FTLD patient brain. Previous studies into DPR protein inclusion frequency have been carried out either by counting inclusions by eye in a small sample size of cells or patients, or in a semi-quantitative manner in which an

estimation of inclusion density is made, often using a four-step scale with categories such as “absent”, “rare”, “moderate” or “frequent”. We therefore sought to develop an unbiased, fully automated, fully quantitative image analysis protocol that can be used to analyse large numbers of cells in order to more accurately assess differences in inclusion frequencies for different DPR proteins.

Immunohistochemical staining for DPR protein and p62 aggregates was optimised and performed by Dr. Tammaryn Lashley at the Queen Square Brain bank on post-mortem brain sections from eight heterozygous *C9orf72* FTLD cases, one homozygous *C9orf72*FTLD case and one *C9orf72* ALS case. Cerebellum, frontal cortex and hippocampus sections from each patient were individually labelled with antibodies against each of the five DPR proteins (poly(GA), poly(GP), poly(GR), poly(AP) and poly(PR)) or p62. These inclusions were then visualised using a brown 3,3’-diaminobenzidine (DAB) chromagen substrate. Cell nuclei were counterstained with the blue nucleic acid stain hematoxylin (Figure 4.7). Due to the absence of discrete channels for each individual antibody signal in this immunohistochemical staining, a more sophisticated image analysis protocol is required than was utilised in the previous chapter to quantify immunofluorescent staining. To achieve this, we collaborated with digital image analyst Matthew Ellis in the Institute of Neurology Division of Neuropathology to design an image analysis ruleset using the software Definiens Developer XD.

The different brain regions stained display variances in cell size, cell density and staining profile, and consequently each area requires a customised analysis protocol. We therefore focussed on the optimisation of an initial DPR analysis protocol in a single brain region, which can be adapted in future for use in other areas. The cerebellum was chosen as the initial analysis region for a number of reasons: 1) the more limited degree of degeneration

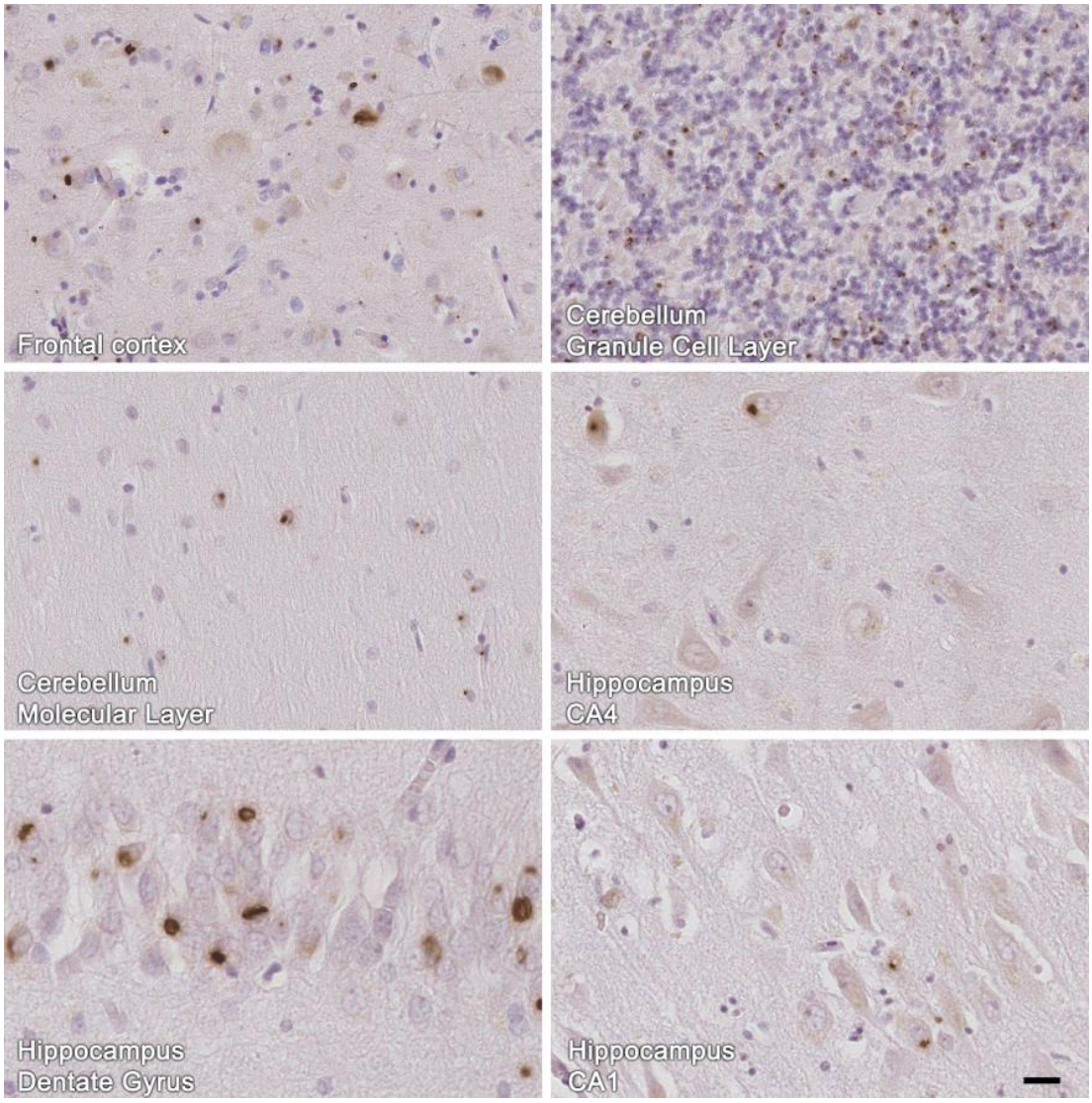


Figure 4.7 p62 inclusions in different regions of *C9orf72* FTD patient brain

Representative images of inclusions in different brain regions from patient #14. Sections were immunostained with p62 primary antibody and visualised with 3,3'-diaminobenzidine (DAB; brown) . Nuclei were visualised with hematoxylin (blue). Scale bar represents 20 µm.

in the cerebellum compared to the frontal cortex facilitates the initial design of a basic analysis protocol, which can then be adapted further to accommodate for the more prevalent debris and tissue damage artefacts present in the frontal cortex, 2) the granule cell layer presents a particular challenge to quantifying inclusions by eye due to the vast number of small, densely packed neurons in this region, and development of an automated protocol would therefore be highly beneficial in this area, 3) examination of DPR

aggregation in the cerebellum may aid understanding of the unique cerebellar pathology observed in *C9orf72* FTD/ALS patients.

Figure 4.8 displays representative images for each of the immunostains to be examined in the cerebellum granule cell layer and molecular layer. Considerable differences exist between the staining profiles of the different antibodies used, with a particularly high background DAB stain observed in slides stained with poly(GA) primary antibody compared to the other stains. The level of DAB background in poly(GA) stained slides also varies

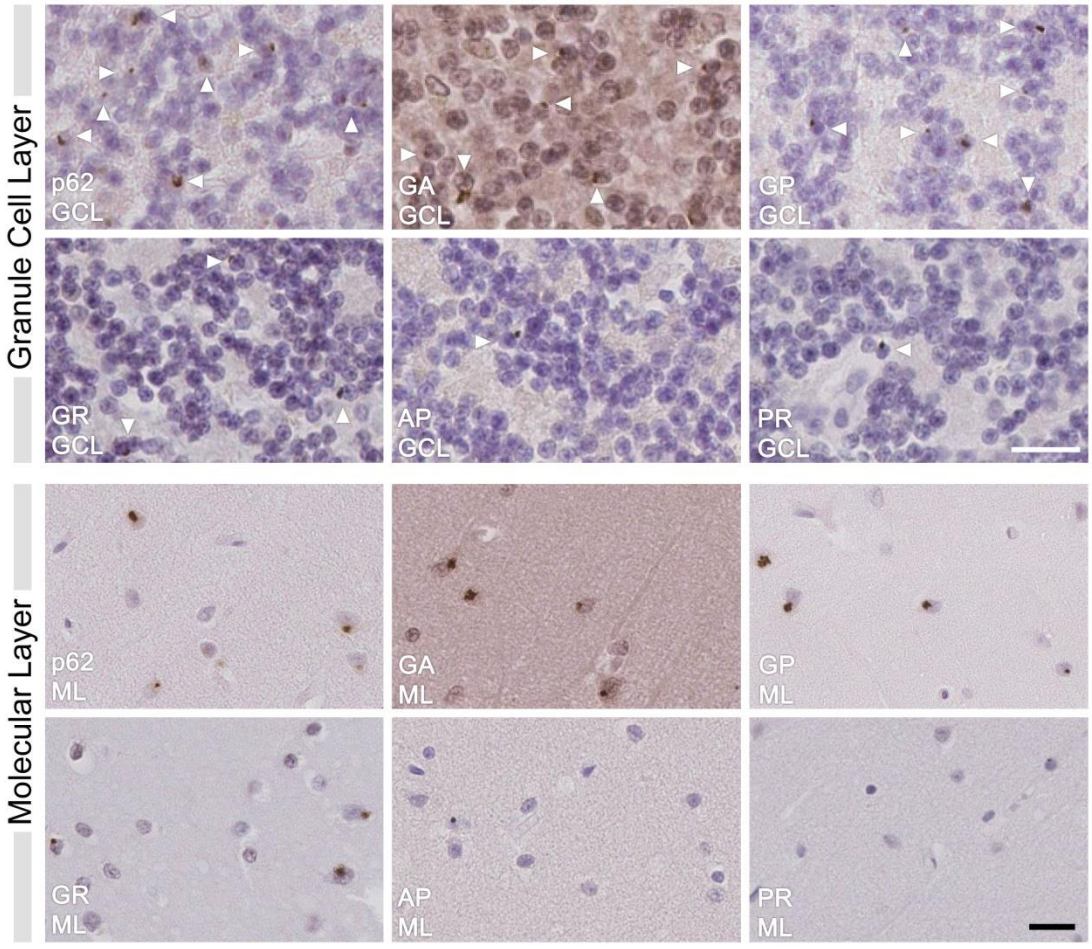


Figure 4.8 p62 and DPR protein inclusions in the cerebellum granule cell layer and molecular layer of *C9orf72* FTLD patients

Representative images of immunostains using different antibodies in the cerebellum granule cell layer (GCL) and molecular layer (ML) including p62, glycine-alanine (GA), glycine-proline (GP), glycine-arginine (GR), alanine-proline (AP), and proline-arginine (PR), all visualised with 3,3'-diaminobenzidine (DAB; brown) . Nuclei were visualised with hematoxylin (blue). White arrows highlight examples of inclusions in the granule cell layer. White scale bar represents 20 µm in the granule cell layer. Black scale bar represents 20 µm in the molecular layer.

greatly between different patients (Figure 4.9). Additionally, poly(GR) stained slides exhibited very dark hematoxylin staining compared to other slides, which reduced the contrast between inclusions and cell nuclei. These inconsistencies necessitated the development of four complimentary analysis rulesets, each customised for a particular staining profile (see methods section 2.5.3 for full details of protocols used). We first developed a basic protocol to identify blue hematoxylin stained nucleli and brown DAB stained inclusions in the majority of stained slides. In this protocol, the threshold for identification of DAB staining is determined using the value for the 95th percentile of brown staining intensity in each region of interest to account for small differences in DAB signal strength between slides. However this relative threshold is prevented from dropping below a set absolute value of 0.4 arbitrary units of colour intensity to avoid false identification of

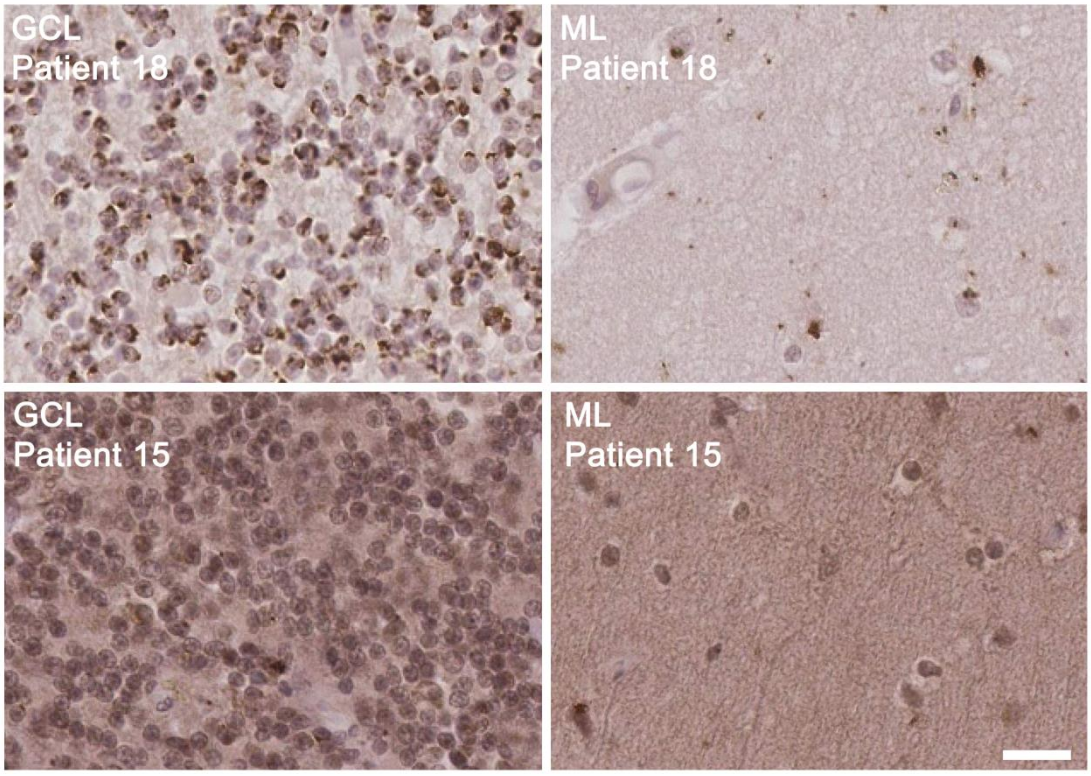


Figure 4.9 Variability of poly(GA) immunostaining in the cerebellum granule cell layer and molecular layer

Representative images of glycine-alanine (GA) immunostaining in the cerebellum granule cell layer (GCL) and molecular layer (ML) in patients #15 and #18, visualised with 3,3'-diaminobenzidine (DAB; brown), highlighting the patient to patient variability in staining. Nuclei were visualised with hematoxylin (blue). Patient #18 was analysed using the 'dark' protocol, patient #15 was analysed using the 'very dark' protocol. Scale bar represents 20 μm.

inclusion objects due to low or absent DAB staining in some specimens. A sophisticated cell splitting algorithm was also designed to isolate individual nuclei from multiple touching cells, which was of particular importance in the densely populated granule cell layer. Within the protocol, two separate subroutines with slight modifications were used to analyse granule cell or molecular layer areas. Identification of cell nuclei in the molecular layer was adapted from that used in the granule cell layer to enable recognition of fainter hematoxylin staining present in these cells. Additionally, due to the presence of frequent blood vessels in the molecular layer, a vessel removal step was included for this area. This step identified areas of blue staining that had a length to width ratio of over 3 or that were over 2000 pixels in size, and excluded them (Figure 4.10).

While this basic algorithm was effective at identifying inclusions in slides stained for p62, poly(GP), poly(AP), and poly(PR) stained slides, it did not accurately characterise inclusions in poly(GA) or poly(GR) stained slides due to the differences in tissue staining. The dark hematoxylin stain in poly(GR) stained specimens resulted in the false classification of some areas of dark purple hematoxylin stained nuclei in the granule cell layer as DAB stained poly(GR) inclusions. We therefore designed an alternate version of the DAB stain identification procedure with an increased minimum threshold for inclusion identification from 0.4 to 0.5 units within the granule cell layer, which counteracted this problem.

The presence of high DAB background immunoreactivity in poly(GA) stained specimens yielded an unacceptable number of objects falsely identified as inclusions. Additionally, the degree of background staining was highly variable between different poly(GA) stained patient specimens, rendering the development of a single suitable protocol for all poly(GA) stained slides problematic (Figure 4.9). We therefore classified these poly(GA) stained slides into two categories “dark” or “very dark” according to the magnitude of background signal as assessed by eye. A “dark” protocol was designed in which the minimum inclusion

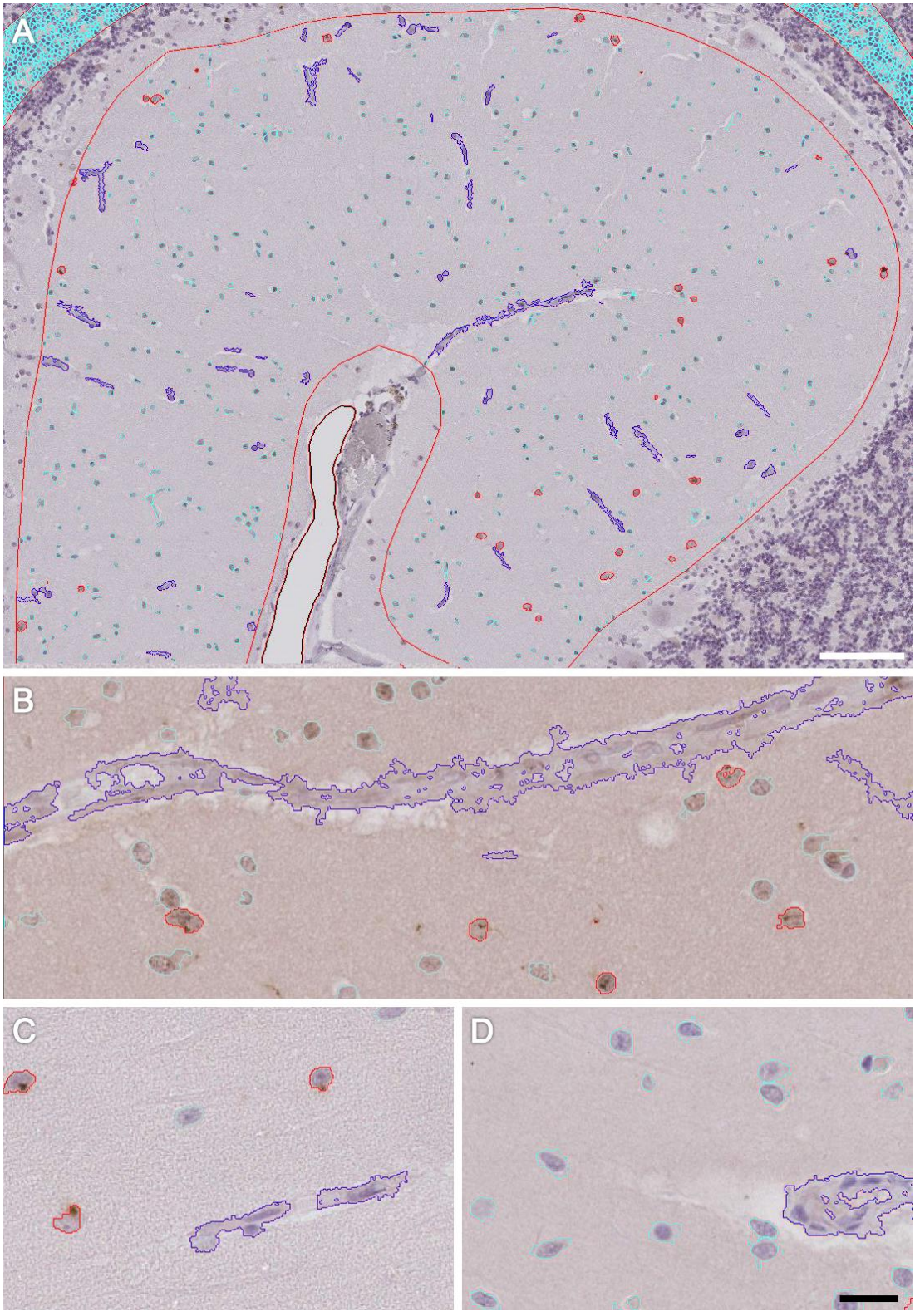


Figure 4.10 Blood vessel identification in the cerebellum molecular layer

Identification and disqualification of blood vessel structures (outlined in dark blue) in the cerebellum molecular layer by the analysis protocol. Inclusion positive cells outlined in red, inclusion negative cells outlined in cyan. A) Lower magnification image of object classification by analysis protocol. White scale bar represents 100 µm. B-D) Higher magnification images of object classification by analysis protocol. Black scale bar represents 20 µm.

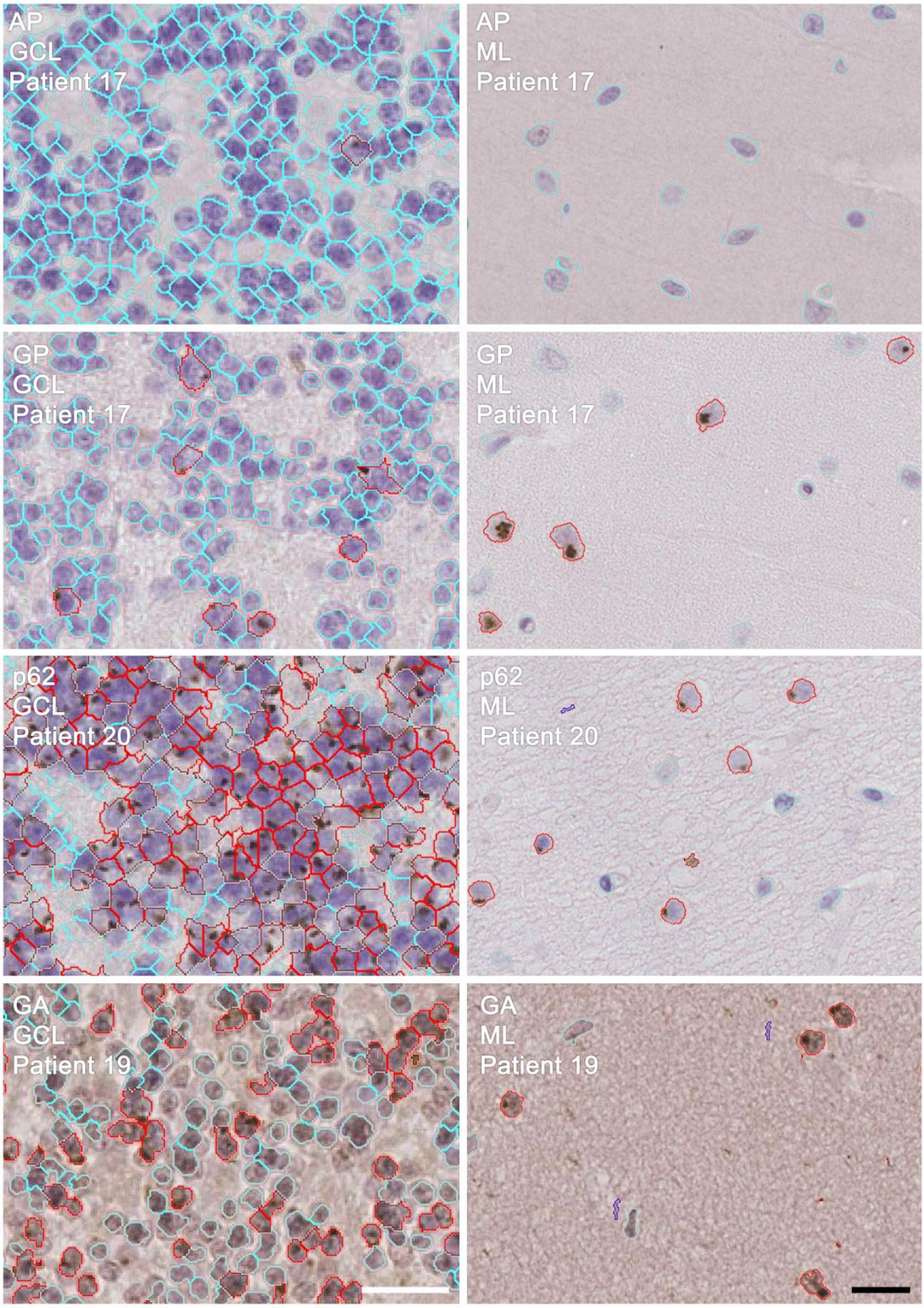


Figure 4.11 Classification of inclusion-positive and inclusion-negative cells by the analysis protocol

Representative images of inclusion-positive (outlined in red) and inclusion negative (outlined in cyan) cells in the cerebellum granule cell layer (GCL) and molecular layer (ML). Images show examples of accurate identification of rare alanine-proline (AP) inclusions, moderate glycine-proline (GP) inclusions, frequent p62 inclusions in the homozygous *C9orf72* FTLD case, and glycine-alanine (GA) inclusions against a high staining background. Primary antibodies were visualised with 3,3'-diaminobenzidine (DAB; brown). Nuclei were visualised with hematoxylin (blue) White scale bar represents 20 µm in the GCL, black scale bar represents 20 µm in the molecular layer.

intensity threshold was increased to 0.7 units, and a “very dark” protocol was designed in which the minimum inclusion intensity threshold was increased to 0.9 units. The initial intensity threshold for inclusion identification was also increased to the 98th percentile for both “dark” and “very dark” protocols. These alterations resulted in a more accurate representation of inclusion burden across different immunostained samples, with fewer regions mistakenly identified as inclusions. Together these adjustments resulted in a toolkit of four protocol versions: normal, dark hematoxylin, dark DAB and very dark DAB. These can be selected depending on the staining profile of each slide to be analysed to yield the most accurate representation of the staining.

The relevant automated analysis protocols were run on the cerebellum granule cell layer and molecular layers for all patients. Figure 4.11 shows examples of the identification of inclusion-positive cells (outlined in red) by the analysis program, in a range of different stains and patients, including rare poly(AP) inclusions, moderate poly(GP) inclusions, frequent p62 inclusions in the homozygous *C9orf72* FTLD case, and poly(GA) inclusions against a high DAB stain background. The resulting values of inclusion-positive and inclusion-negative cells in each area examined were exported for analysis.

4.2.5 DPR protein inclusion frequency in the cerebellum granule cell layer

The percentage of cells within the cerebellar granule cell layer that are positive for each of the five DPR protein inclusions and p62 was plotted for each patient (Figure 4.12). p62 and poly(GA) positive inclusions were found to be most frequent, present in a mean of 11.94 ± 5.06 % and 13.39 ± 4.73 % of cells respectively. Poly(GP) inclusions were approximately half as frequent as p62 or poly(GA) inclusions, with an average of 5.21 ± 3.50 % of cells found to be positive. Poly(GR) inclusions were infrequent in this brain area, with 1.36 ± 0.68 % of cells exhibiting these inclusions on average. Lastly, the antisense repeat transcript products poly(AP) and poly(PR) were present in the lowest percentage of cells, with a mean of 0.11 ±

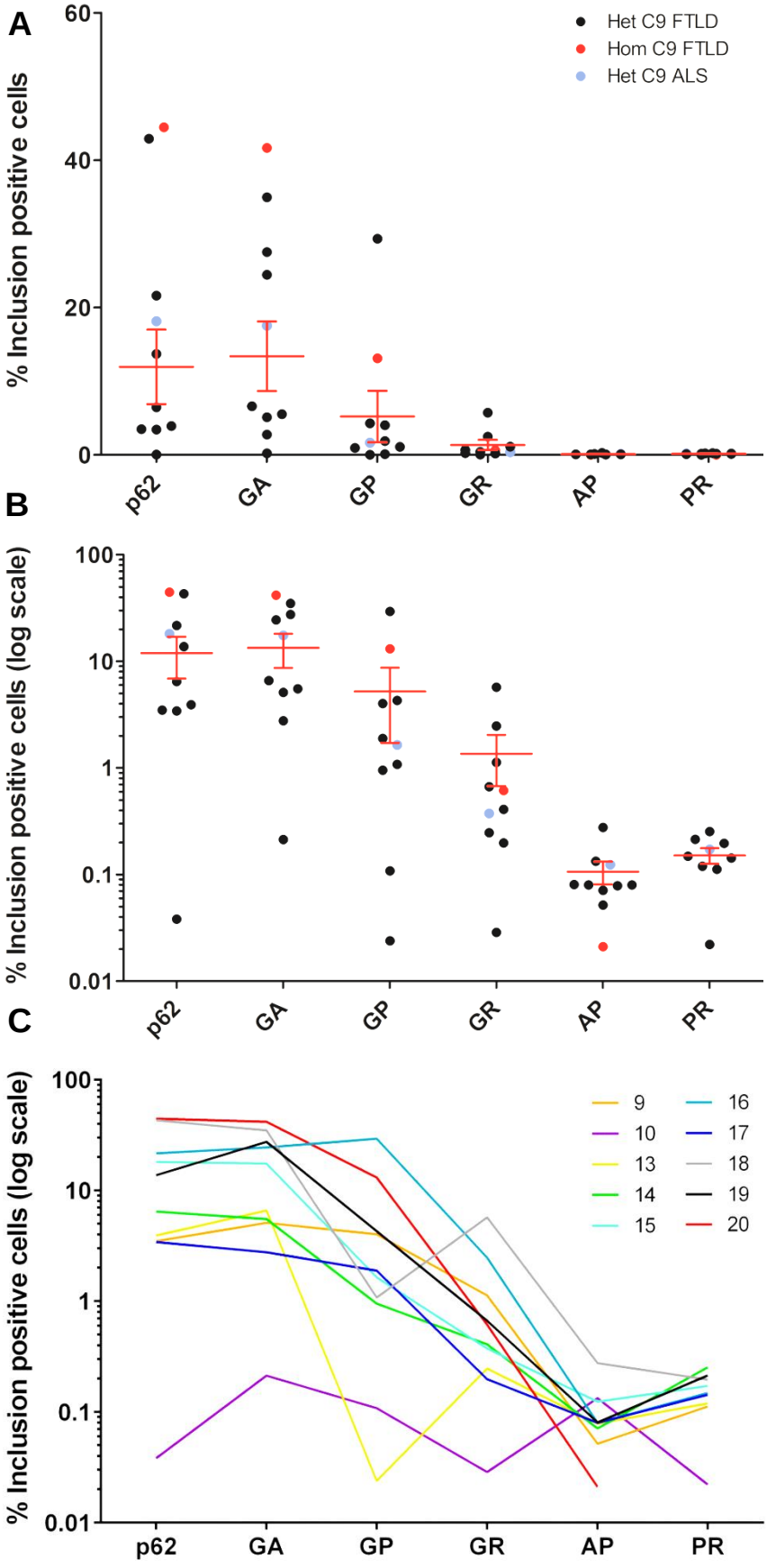


Figure 4.12 Quantification of DPR protein inclusion burden in the cerebellum granule cell layer.

A-B) The percentage of cells positive for glycine-alanine (GA), glycine-proline (GP), glycine-arginine (GR), alanine-proline (AP), and proline-arginine (PR) DPR protein or p62 inclusions in the granule cell layer was plotted for each patient analysed. Red horizontal line represents the mean inclusion frequency in heterozygous (het) FTLD patients. Error bars represent the standard error of the mean (SEM). Red points indicate homozygous (hom) FTLD patient. Blue points indicate het ALS patient. points representing zero positive cells were omitted from the logarithmic scale graph.

C) The percentages of inclusion positive cells for each inclusion type were connected for each patient to show variation of inclusion burden within each case.

Legend displays patient ID numbers next to the relevant line colour. Points representing zero positive cells were omitted from the logarithmic scale graph.

0.03 % and 0.15 ± 0.03 % of cells found to be positive for these inclusions. A high degree of variation between patients was observed in the percentage of cells exhibiting a given inclusion, particularly between the levels of p62, poly(GA) and poly(GR) inclusions with standard deviations of 14.3, 13.4 and 9.9 % observed respectively. As a result of this high variance, no statistically significant difference was detected between the means of any of the inclusion frequencies in heterozygous *C9orf72* FTLD patients by one-way ANOVA when all pairs of individual group means were compared by Bonferroni's multiple comparison test.

A number of outliers were observed in this data set. These were examined by eye, and the accuracy of the analysis protocol quantification was confirmed in each case. Heterozygous *C9orf72* FTLD patient #10 was observed to consistently exhibit very few inclusions of any DPR protein in the cerebellum granule cell layer, despite robust DPR pathology in other brain areas including the adjacent molecular layer, with inclusions observed in a minimum of 0.02 % of cells for poly(PR) inclusions, to a maximum of 0.21 % of cells for poly(GA) inclusions. Heterozygous *C9orf72* FTLD patient #13 was also observed to exhibit an unusually low percentage of poly(GP) positive cells (0.02 % of cells), while the number of inclusions of other DPR proteins and p62 fell closer to the group means.

A single homozygous *C9orf72* FTLD patient (#20) and a single heterozygous *C9orf72* ALS patient (#15) were also tested in this data set, but were not included in the statistical analyses performed. The homozygous *C9orf72* FTLD patient #20 exhibited the highest percentage of p62 and poly(GA) inclusion positive cells of all patients tested (44.5 and 41.1 % of cells respectively), and the second highest percentage of poly(GP) positive cells (13.1 % of cells), however this patient also exhibited a below average burden of poly(AP) and poly(PR) inclusions (0.02 and 0 % of cells within the area analysed respectively). The

heterozygous *C9orf72* ALS case #15 did not show any clear deviation from the heterozygous FTLD mean frequency of any of the inclusion proteins tested.

4.2.6 DPR protein inclusion frequency in the cerebellum molecular layer

The percentage of DPR protein or p62 inclusion positive cells within the cerebellar molecular layer was plotted for each patient (Figure 4.13). In this region, the analysis protocol detected poly(GA) inclusions in a higher proportion of cells than any other marker, with a mean of 22.07 ± 3.95 % of cells containing an inclusion. p62, poly(GP) and poly(GR) positive inclusions were detected in approximately one third as many cells as poly(GA), with an average of 8.97 ± 2.76 %, 7.96 ± 2.32 % and 6.94 ± 1.67 % of cells respectively found to contain these inclusions. Finally, poly(AP) and poly(PR) inclusions were once more found to be the least frequent in this area, with 0.35 ± 0.09 % and 0.17 ± 0.06 % of cells respectively found to be positive on average. A high degree of patient to patient variability was again observed in the percentage of cells exhibiting any given inclusion, with a highest standard deviation of 11.2 % observed for poly(GA) inclusions. When a one-way ANOVA was performed on the data set and all pairs of inclusion group means were compared, a significant difference was observed between the frequency of poly(GA) inclusions and all other inclusions (corrected p values: p62 $p < 0.01$, poly(GP) $p < 0.01$, poly(GR) $p < 0.001$, poly(AP) $p < 0.0001$, poly(PR) $p < 0.0001$), however no significant difference was observed between any other pairs tested.

In contrast to observations in the granular cell layer, the heterozygous *C9orf72* FTLD patient #10 exhibited robust DPR pathology within the cerebellum molecular layer. However, similar to observations in the granule cell layer, a very low number of poly(GP) inclusions were detected in the cerebellum molecular layer of heterozygous *C9orf72* FTLD patient #13, with only 0.15 % of cells exhibiting an inclusion, while levels of other DPR proteins and p62 in this patient once again fell closer to the group means.

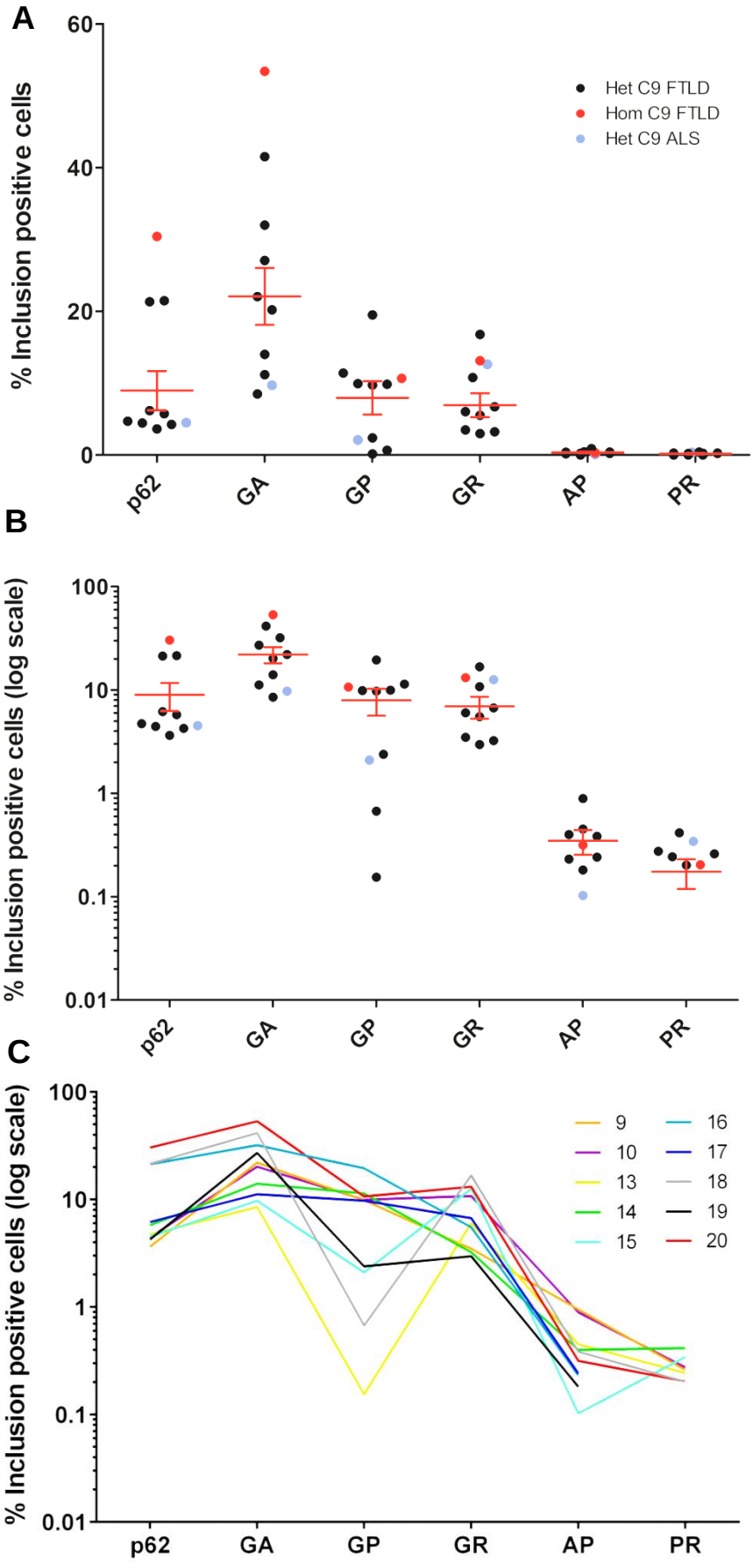


Figure 4.13 Quantification of DPR protein inclusion burden in the cerebellum molecular layer.

A-B) The percentage of cells positive for glycine-alanine (GA), glycine-proline (GP), glycine-arginine (GR), alanine-proline (AP), and proline-arginine (PR) DPR protein or p62 inclusions in the molecular layer was plotted for each patient analysed. Red horizontal line represents the mean inclusion frequency in heterozygous (het) FTLD patients. Error bars represent the standard error of the mean (SEM). Red points indicate homozygous (hom) FTLD patient. Blue points indicate het ALS patient. points representing zero positive cells were omitted from the logarithmic scale graph.

C) The percentages of inclusion positive cells for each inclusion type were connected for each patient to show variation of inclusion burden within each case.

Legend displays patient ID numbers next to the relevant line colour. Points representing zero positive cells were omitted from the logarithmic scale graph.

The homozygous *C9orf72* FTLD patient #20 exhibited the highest percentage of p62 and poly(GA) inclusion positive cells of all patients tested (30.4 and 53.4 % of cells respectively) as was observed in the granule cell layer, but fell closer to the heterozygous *C9orf72* FTD group means with regards to the frequency of poly(GP), poly(GR), poly(AP), and poly(PR) inclusions. The heterozygous *C9orf72* ALS case #15 did not show any clear deviation from the heterozygous FTLD mean frequency of any of the inclusion proteins tested in this region, as was also observed in the granule cell layer.

4.2.7 Comparison of DPR protein expression in the cerebellum granule cell layer and molecular layer

We next looked to compare the DPR protein inclusion burden in the cerebellum molecular layer and granule cell layers. Firstly, the frequency of each type of inclusion in the granule cell layer was plotted against the corresponding frequency in the molecular layer in each patient (Figure 4.14). Linear regressions were carried out for each inclusion type, and a significantly non-zero slope was found for p62 ($p < 0.001$, $r^2 = 0.76$) and poly(GA) ($p < 0.001$, $r^2 = 0.74$) inclusion frequency, indicating that a patient with a high burden of these inclusions in one of the regions is likely to have a high burden in the other. A marginally significant correlation was also observed for poly(GP) inclusion frequency ($p = 0.494$, $r^2 = 0.50$), and a marginally non-significant correlation was observed for poly(GR) inclusion frequency ($p = 0.0533$, $r^2 = 0.49$). However, no significant correlation was observed between the regions for poly(AP) or poly(PR) inclusion frequency.

Next, we examined whether there was a difference in inclusion burden between the two areas for any of the inclusion types examined. To do this, the percentage of inclusion-positive cells in the molecular layer was divided by the percentage of inclusion-positive cells in the granule cell layer for each inclusion type in each patient to give the fold change between the two areas (Figure 4.15). Plotting these values highlights the regional disparity

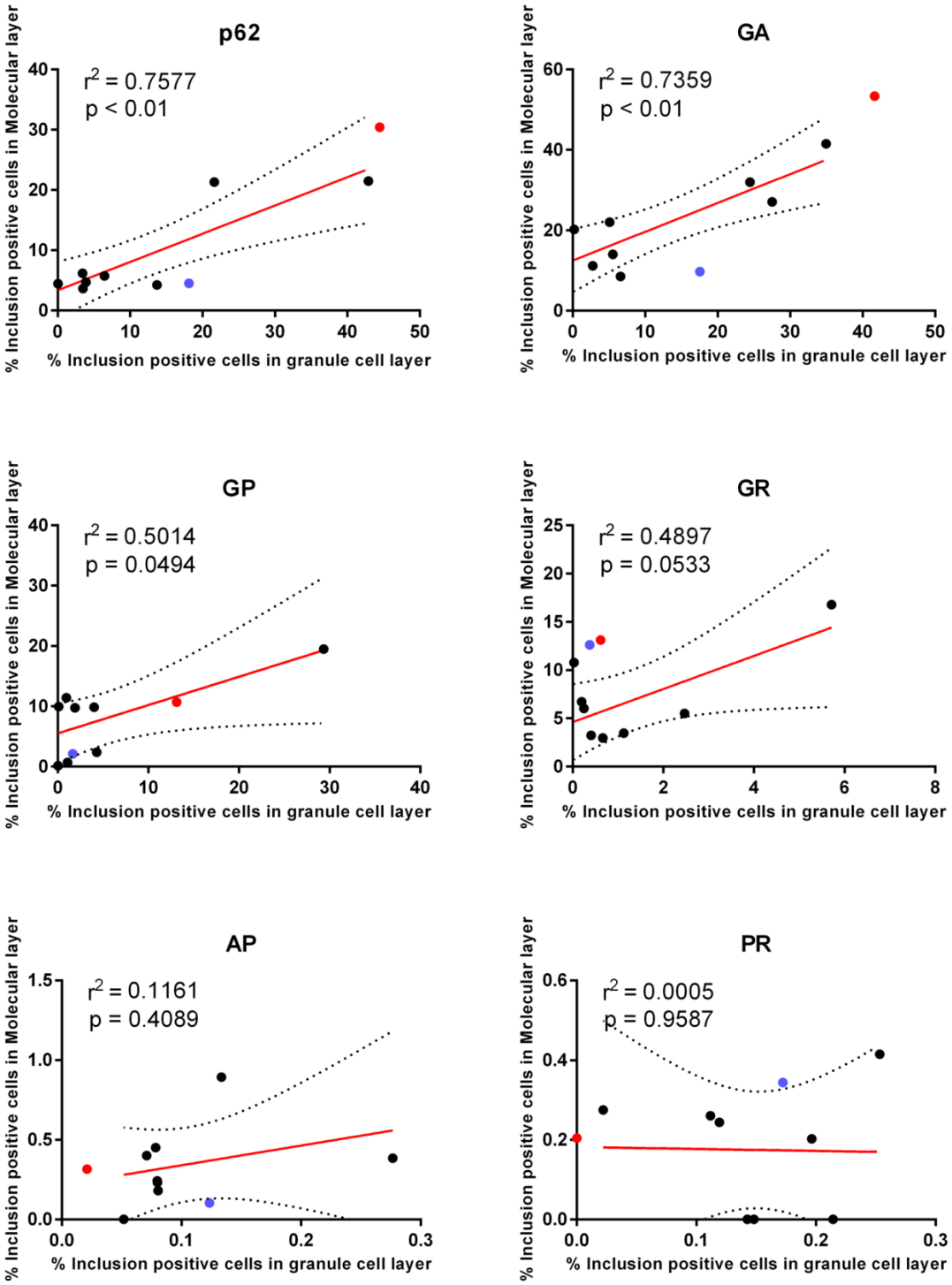


Figure 4.14 Correlation of inclusion burden in the cerebellum molecular layer and granule cell layers

The percentage of inclusion-positive cells in the cerebellum granule cell layer was plotted against the percentage of inclusion-positive cells in the cerebellum molecular layer for p62, glycine-alanine (GA), glycine-proline (GP), glycine-arginine (GR), alanine-proline (AP), and proline-arginine (PR) inclusions. Red points indicate homozygous *C9*orf72 FTLD patient, blue points represent heterozygous *C9*orf72 ALS patient. Linear regressions were carried out on heterozygous FTLD patients for each comparison. Red lines represent lines of best fit. Dashed black lines show 95 % confidence band for line of best fit. P values shown denote the statistical probability of a linear trend with a non-zero slope. r^2 values are shown as a measure of goodness of fit.

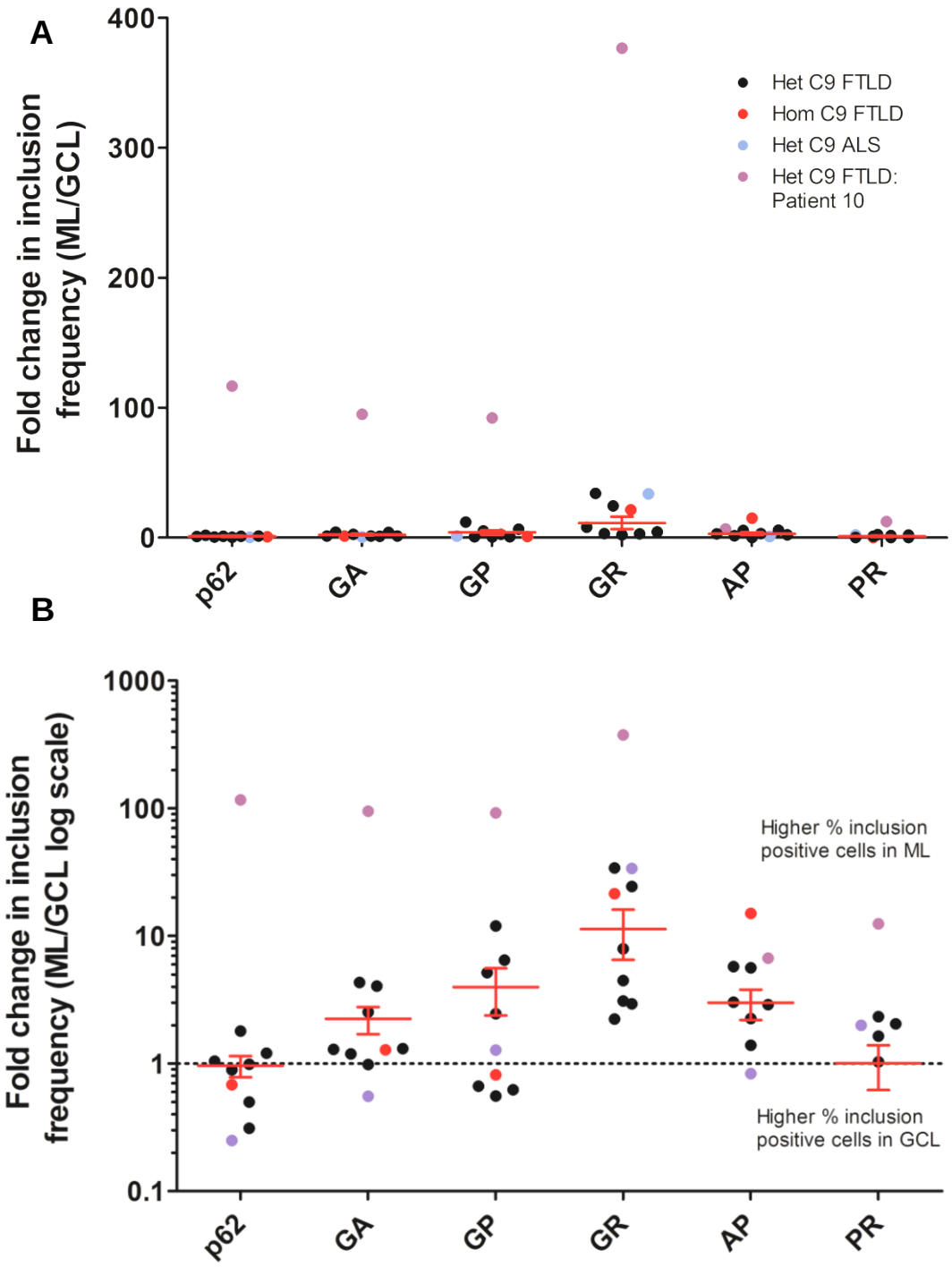


Figure 4.15 Comparison of inclusion frequency in the cerebellum granule cell and molecular layers

For each patient, the percentage of inclusion positive cells in the molecular layer was divided by the percentage of inclusion positive cells in the granule cell layer for p62, glycine-alanine (GA), glycine-proline (GP), glycine-arginine (GR), alanine-proline (AP), and proline-arginine (PR) inclusions. Red points indicate homozygous *C9*orf72 FTLD patient, blue points represent heterozygous *C9*orf72 ALS patient, purple points represent outlier patient #10. Points were omitted when either of the molecular layer or granule cell layer had zero positive cells for a given inclusion type. Dotted line represents boundary at which there is no change in inclusion frequency between the two regions.

in inclusion burden exhibited by heterozygous *C9orf72* FTLD patient #10, which exhibited 117, 95, 92 and 377 times more inclusions of p62, poy(GA), poy(GP), and poly(GR) respectively in the molecular layer compared to the granule cell layer. This patient was subsequently excluded from the means and standard errors plotted of the remaining heterozygous *C9orf72* FTLD patients to avoid biasing these values to this outlying point. In the remaining patients, an 11.3 ± 4.8 fold higher percentage of poly(GR) inclusion positive cells was observed on average in the molecular layer compared to the granule cell layer. Approximately double the percentage of poly(GA) inclusion-bearing cells (2.2 ± 0.5 fold), and triple the poly(AP) inclusion-bearing cells (3.0 ± 0.8 fold) was also observed in the molecular layer compared to the granule cell layer, while p62 and poly(PR) burden was similar in both regions.

4.2.8 Examining relationship between *C9orf72* disease course and DPR burden in the cerebellum

We next examined whether any relationship could be observed between the inclusion burden of any of the DPR proteins or p62 within the cerebellum molecular layer or granule cell layer, and the disease course of the patients studied (Figure 4.16). The percentage of cells positive for each inclusion type in either the molecular layer or granule cell layer was plotted against either patient age of disease onset, or disease duration. However, no significant correlation was found for any of the comparisons tested.

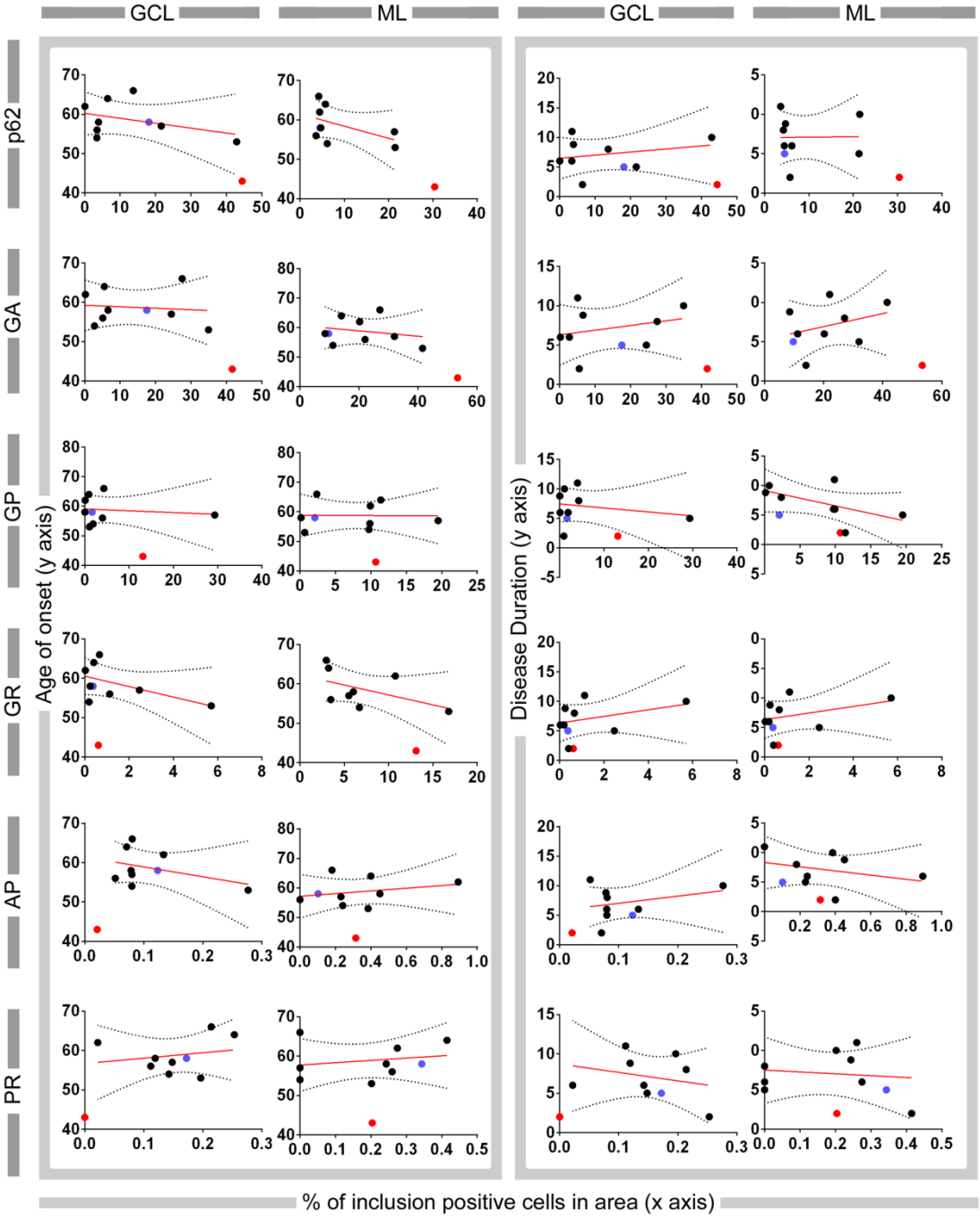


Figure 4.16 Correlation of inclusion burden with patient disease course

The percentage of inclusion-positive cells for p62, glycine-alanine (GA), glycine-proline (GP), glycine-arginine (GR), alanine-proline (AP), and proline-arginine (PR) inclusions in the cerebellum granule cell layer (GCL) or molecular layer (ML) was plotted against patient age of onset or disease duration. Red points indicate homozygous *C9orf72* FTLD patient, blue points represent heterozygous *C9orf72* ALS patient. Linear regressions were carried out on heterozygous FTLD patients for each comparison. Red lines represent lines of best fit. Dashed black lines show 95 % confidence band for line of best fit.

4.3 Discussion

4.3.1 Summary of results

In this chapter, we investigated a role for DPR protein gain-of-function disease mechanisms in *C9orf72* FTD/ALS by examining the consequences of DPR protein production in a cell model system, in transgenic *Drosophila*, and in human post-mortem brain.

We expanded our toolkit of DNA constructs to enable the dissection RNA gain-of-function from DPR protein gain-of-function mechanisms by generating a set of 'protein-only' repeat constructs that individually code for each of the five DPR proteins, but do not produce repetitive RNA. When transiently transfected into a HEK293T cell model system, $(GA)_{100}$ formed star-like inclusions similar to those observed in human *C9orf72* FTD/ALS patient post-mortem brain. $(PR)_{36}$ and GFP tagged $(GR)_6$-GFP and $(PR)_6$-GFP constructs expressed in these cells were observed to aggregate within the nucleus, mirroring the nucleolar localisation profile observed by other groups. Conversely, in longer $(PR)_{100}$ constructs, aggregates were observed to form in the cytoplasm, suggesting that repeat length can affect DPR protein localisation.

We then expressed these protein-only constructs in a *Drosophila melanogaster* model system, and found that expression of the arginine-containing DPR proteins poly(GR) and poly(PR) was sufficient to cause severe degeneration when expressed in the fly eye or adult neurons, implicating a toxic role for these proteins in *C9orf72* pathology. A mild reduction in viability was caused by the expression of poly(GA) repeats in fly adult neurons, but no degeneration was observed as a result of poly(AP) expression.

Finally, we developed an automated, high throughput and unbiased protocol for quantifying DPR protein and p62 aggregation in immunohistochemically stained sections of *C9orf72* FTD/ALS patient post-mortem brain, to enable examination of the relevance of

DPR protein toxicity to disease in human patients. An array of complimentary rulesets were developed for use with several differing staining profiles, which were then applied to stained patient sections of the cerebellum granule cell and molecular layers. p62 and poly(GA) positive inclusions were found to be the most frequent, while fewer poly(GP) and poly(GR) inclusions were observed, and poly(AP) and poly(PR) aggregates were detected only rarely. A disparity in inclusion frequency was observed between the cerebellum molecular layer and granule cell layers, with a trend for higher poly(GR) inclusion burden in the molecular layer observed. Furthermore, DPR protein inclusions were found to be almost absent in the granule cell layer of one FTLD patient who had robust DPR pathology in the molecular layer. Finally, we examined whether inclusion burden in the cerebellum molecular layer or granule cell layer correlated with patient disease course, but found no correlation.

4.3.2 Subcellular localisation of DPR proteins in cell models

To study the subcellular localisation of the different DPR proteins, our protein-only repeat constructs were transiently transfected into a HEK293T cell line. While this cell line was primarily chosen due to its high transfection efficiency rate and its wide use in the study of DPR proteins and RAN translation in *C9orf72* FTD/ALS, it has also been found that HEK293 cells expresses a number of markers that are associated with early differentiating neuronal cells (Shaw et al., 2002), suggesting that this cell line does not exclusively represent a kidney epithelial lineage.

The subcellular localisation of DPR proteins observed in HEK293T cells transfected with our constructs is largely consistent with studies of constructs and cell models by other groups, however some interesting differences in subcellular localisation can be observed between different reports, raising the question of which factors might affect the localisation of these proteins. Several other groups have reported that cells transfected with poly(GA)

constructs form star-like cytoplasmic aggregates similar to those seen in patient brain (May et al., 2014; Schludi et al., 2015; Wen et al., 2014; Yamakawa et al., 2014; Yang et al., 2015), while others have also observed diffuse, cytoplasmic poly(GA) signal (Tao et al., 2015; Yamakawa et al., 2014; Zu et al., 2013). Increasing repeat length was observed to promote more extensive poly(GA) aggregate formation in our model, however repeat length alone does not predict the presence of diffuse or aggregating poly(GA) as both forms have been observed in cells transfected with long constructs with over 100 repeats as well as in constructs as short as 30 repeats (May et al., 2014; Schludi et al., 2015; Tao et al., 2015; Wen et al., 2014; Yamakawa et al., 2014; Yang et al., 2015; Zu et al., 2013). This difference also does not appear to be exclusively driven by cell type, as both diffuse and aggregate forming poly(GA) has been observed in HEK293T cells (May et al., 2014; Tao et al., 2015; Yamakawa et al., 2014; Zu et al., 2013). In our cell system, insoluble poly(GA) aggregates were recognised, however soluble poly(GA) that exhibited a diffuse, cytoplasmic localisation in our $(GA)_6$-GFP tagged constructs was not bound by our antibody. This suggests that soluble or insoluble poly(GA) repeats may be differentially recognised depending on the antibody used, and that both soluble and insoluble forms of poly(GA) are most likely present in transfected cells.

In our model, $(GR)_6$-GFP and $(PR)_6$-GFP constructs were observed to aggregate in the nucleus, recapitulating the nucleolar localisation observed in most cell models (Kwon et al., 2014; May et al., 2014; Schludi et al., 2015; Tao et al., 2015; Wen et al., 2014; Yamakawa et al., 2014; Zu et al., 2013). An interesting difference in the subcellular localisation of poly(PR) aggregates was observed between our PR_{36} and PR_{100} transfected cells. In PR_{36} transfected cells, anti-poly(PR) antibody signal primarily detected intranuclear aggregates, while in PR_{100} transfected cells the inclusions were primarily cytoplasmic. This suggests that poly(PR) may be differentially processed by cells depending on peptide length. Recent work has shown that nucleocytoplasmic transport is disrupted in models of *C9orf72* FTD/ALS

(Freibaum et al., 2015; Jovičić et al., 2015; Zhang et al., 2015), with $(PR)_{50}$ treatment specifically found to disrupt nuclear protein import in yeast (Jovičić et al., 2015). Longer poly(PR) repeat length may therefore disrupt nuclear import and disfavour nucleolar binding, promoting the formation of cytoplasmic aggregates. However, cellular models with a range of different length poly(PR) repeats have been observed to form both intranuclear and cytoplasmic poly(PR) aggregates, indicating that once again multiple different factors such as cell type, repeat length and antibody used determine the observed subcellular localisation of poly(PR) (May et al., 2014; Schludi et al., 2015; Tao et al., 2015; Wen et al., 2014; Yamakawa et al., 2014; Zu et al., 2013).

In our cell model GFP signal could be detected within the nucleus of cells transfected with all $(DPR)_{6/7}$-GFP constructs, however the anti-DPR protein antibodies used detected only cytoplasmic protein. This could reflect an inability for the antibody to enter the nucleus due to insufficient permeabilisation, however, the detection of intranuclear poly(PR) aggregates in PR_{36} and PR_{100} transfected cells suggests that this is not the case. It may therefore be beneficial to generate untagged $(GR)_6$ and $(PR)_6$ constructs to determine whether the presence of the tag is affecting antibody recognition of these DPR protein epitopes.

Both poly(GP) and poly(AP) have consistently been detected in a diffuse cytoplasmic localisation in cell models by numerous studies (May et al., 2014; Schludi et al., 2015; Tao et al., 2015; Wen et al., 2014; Yamakawa et al., 2014; Zu et al., 2013). An interesting aspect to studying poly(GP) production in cells is that this is the only DPR protein produced from both sense and antisense transcripts. One group developed poly(GP) protein-only constructs that additionally mimic the unique downstream regions of the respective sense and antisense translated poly(GP) frames, however both constructs were observed to exhibit similar diffuse cytoplasmic localisation in transfected cells (Zu et al., 2013). While our $(GP)_6$-GFP, $(AP)_7$-GFP, AP_{36} and AP_{100} constructs all exhibited cytoplasmic DPR signal, no

signal could be detected in our GP_{36} transfected cells. As we were unable to generate a GP_{100} protein-only construct, it is not known whether longer poly(GP) proteins can be detected in our system. It may be beneficial to attempt to generate an intermediate length poly(GP) construct between 36 and 100 repeats in order to determine whether increased repeat length would enable the detection of this DPR protein in our system.

4.3.3 DPR protein expression in a transgenic *Drosophila* model

In the previous chapter, we found that while pure constructs that produce both GGGGCC repeat RNA and DPR proteins caused degeneration in a *Drosophila* model, RNA-only constructs which produce repetitive RNA but not DPR proteins did not. This implicated DPR proteins as the species responsible for toxicity in this model. Treatment of flies expressing 36 or 103 pure repeats with a non-lethal dose of cyclohexamide, an inhibitor of protein biosynthesis, was found to alleviate the reduction in lifespan caused by these constructs, further supporting a protein gain-of-function mechanism of toxicity (Mizielinska et al., 2014). To further investigate this possibility, we produced flies expressing protein-only constructs coding for poly(GA), poly(GR), poly(PR) and poly(AP) DPR proteins of 36 and 100 repeats in length, to determine which of the individual DPR proteins are sufficient to cause a toxic effect. The arginine containing DPR proteins poly(GR) and poly(PR) were found to cause severe degeneration when expressed in the *Drosophila* eye and an abrupt drop in viability when expression was induced in adult neurons, which was more severe in flies expressing 100 protein-only repeats compared to 36 repeats. In contrast, no effect was seen in poly(AP) expressing flies, and only a modest decrease in survival was caused by induction of poly(GA) expression in adult neurons, with no visible degeneration observed when poly(GA) was expressed in fly eye. Arginine-containing DPR proteins are therefore sufficient to cause toxicity in this *Drosophila* model, implicating that these proteins could play an important role in *C9orf72* FTD/ALS pathology.

Evidence for toxicity by poly(GR) and poly(PR) arginine-containing DPR proteins has also been reported in numerous other models. Several other transgenic *Drosophila* models have been generated subsequent to the publication of our model, which also exhibit strong toxicity upon expression of poly(GR) and poly(PR) protein-only repeat constructs, but not upon expression of other DPR proteins (Wen et al., 2014; Yang et al., 2015). One study also found that expression of 160 pure repeats within the first intron of an artificial *C9orf72* minigene in *Drosophila* eye or adult neurons did not exhibit an obvious neurodegenerative phenotype, which coincided with a lack of nuclear export of repeat RNA from this construct, and therefore reduced RAN translation of DPR protein products in the cytoplasm (Tran et al., 2015). Transformation of PR_{50} into a *Saccharomyces cerevisiae* yeast cell model also exhibited toxicity, which was found to be modified by numerous factors involved in nucleocytoplasmic transport, indicating that this process could be a mechanism through which arginine-containing DPR protein toxicity operates (Jovičić et al., 2015).

4.3.4 Evidence for arginine containing DPR protein toxicity in cell culture models

Poly(GR) and poly(PR) induced toxicity has also been observed in a number of cell culture models. Exogenously applied synthetic GR_{20} and PR_{20} DPR proteins were observed to bind to nucleoli and induce toxicity in cultured human astrocytes and U2OS cells. Furthermore, cells treated with PR_{20} exhibited aberrant RNA processing, including missplicing and reduced maturation of rRNA (Kwon et al., 2014). Transient transfection of PR_{50} and GR_{50} protein-only constructs also reduced survival in cultured rat primary cortical and motor neurons. Furthermore, longitudinal live-cell imaging study of PR_{50}-GFP transfected cortical neurons found that PR_{50}-GFP aggregate bearing neurons exhibited a significantly higher risk of death compared to transfected neurons with diffuse PR_{50}-GFP signal. Toxicity was also observed in cortical neurons transfected with poly(PR) constructs as short as 25 repeats, at which length poly(GR) was not observed to be toxic, suggesting that poly(PR) may be the more potent toxic species (Wen et al., 2014). Together, these studies support a pathogenic

role for poly(PR) and poly(GR), and additionally implicate that disruption of nucleolar function as a mechanism toxicity

Arginine-containing DPR proteins have additionally been observed to cause toxicity in some transiently transfected immortalised cell lines (Tao et al., 2015; Yamakawa et al., 2014), but not in others (May et al., 2014). Increased cell death was detected by propidium iodide staining in NSC-34 and HEK293 cell lines transfected with GFP tagged GR_{30} and PR_{30} protein-only constructs, with no toxicity observed in cells transfected with other DPR protein constructs of a similar length compared to GFP only controls (Tao et al., 2015). Additionally, transfection of Neuro2a cells with GR_{100} protein-only constructs increased caspase-3 immunoreactivity compared to GFP transfected controls, however this toxicity was not observed in PR_{100} transfected cells (Yamakawa et al., 2014). Another study investigated toxicity using an LDH assay in HEK293 cells transfected with protein-only constructs between 80 and 175 repeats in length, but did not observe toxicity for any of the five DPR proteins tested (May et al., 2014). The variable toxicity exerted by poly(GR) and poly(PR) constructs in different model systems highlights that numerous environmental factors such as repeat length, cell type, incubation time, expression level and cellular stressors may modulate the action of these molecules. The context of poly(GR) and poly(PR) expression in patient brain may therefore also be important in modulating the extent to which these molecules cause toxicity.

4.3.5 Rationale for use of an automated image analysis protocol in determining DPR protein burden in *C9orf72* patient post-mortem brain

A number of studies have been carried out examining the distribution of DPR protein inclusions in *C9orf72* patient brain, however DPR proteins aggregates have thus far been found not to correlate with the degree of degeneration in different brain regions or with TDP-43 deposition (Davidson et al., 2014; Gomez-Deza et al., 2015; Mackenzie et al., 2013;

Schludi et al., 2015), despite the strong toxicity shown by arginine-containing dipeptides in cell and *Drosophila* models. However, other research has found that DPR protein deposition predates TDP-43 inclusion formation (Baborie et al., 2014; Proudfoot et al., 2014). The global extent of poly(GA) aggregation has also been observed to negatively correlate with disease onset (Davidson et al., 2014), and several conflicting studies have been released suggesting that different DPR protein inclusion profiles might be attributed to different disease subtypes (Gendron et al., 2015; Schludi et al., 2015). A more in-depth quantitative analysis of how DPR inclusion distribution profile differs between brain regions and between different patients may therefore help to determine how these entities might modulate the disease.

In this chapter, we designed and tested an automated analysis protocol for determining the frequency of different DPR protein and p62 inclusions in *C9orf72* post mortem brain. Analysis of inclusion frequency is commonly carried out in a semi-quantitative manner in which inclusion frequency within a region is scored in categories such as “absent”, “rare”, “moderate” or “frequent”. Alternately, inclusions can be counted exactly by eye. However, these methods may be subject to bias or inconsistency: a given field of cells may be counted or classified differently by different researchers, or indeed by the same researcher on a different day depending on the context in which the field is viewed. Furthermore, the time consuming nature of counting by eye limits the sample size of cells, patients and brain regions that can be accurately assessed. Conversely, automated image analysis allows inclusion burden to be determined in a large number of cells and patients in a consistent, unbiased manner. Assessment of inclusions in a large sample size increases the statistical power of the analysis, enhancing the ability to detect differences between different patients and inclusion types.

To detect DPR inclusions, we employed an immunohistochemical staining method using 3,3'-diaminobenzidine (DAB) and hematoxylin, which affords a high sensitivity of epitope detection in comparison to immunofluorescent staining due to greater signal amplification. However, automated analysis of immunohistochemistry with DAB and hematoxylin presents additional difficulties compared to immunofluorescence due to the absence of discrete channels for each antibody signal. The exact hue and intensity of each stain as well as the intensity of background signal can vary considerably between different antibodies, brain regions and quality of patient tissue, necessitating the use of complex computational algorithms for stain recognition (van Der Laak et al., 2000). Additionally, differences can arise between samples if the entire cohort to be analysed is not stained at the same time, as is often the case when large numbers of sections are to be processed. This inherent variation must therefore be accounted for in any automated analysis protocol used. In order to detect inclusion frequencies across all specimens with a similar degree of accuracy, areas of staining were identified by determining relative intensity compared to background staining levels. For specimens that differed substantially from the norm, such as the poly(GA) and poly(GR) sections examined, modified protocols were designed that were optimised for accurate representation these different staining profiles. Ideally, immunostaining should be optimised to maximise consistency between stains and enable analysis with a single all-encompassing protocol, however this is not always possible with the antibodies available. In this study, we manually selected the most suitable protocol for use with each specimen, however the program could be adapted in future to calculate the profile of DAB or hemotoxylin stain intensity exhibited in slides to be analysed and automatically run the most suitable ruleset in order to completely remove any user bias and perform a fully automated and impartial analysis.

4.3.6 Relative frequencies of DPR proteins in *C9orf72* FTD/ALS patient brain

The quantification of DPR protein frequency performed in this chapter suggests that different DPR protein inclusions are present in the brain at very different frequencies, with poly(GA) inclusions found to be the most frequent, poly(GP) and poly(GR) inclusions less frequent, and antisense transcript RAN translation products poly(AP) and poly(PR) inclusions observed only rarely, as has been previously been reported (Mackenzie et al., 2015; Schludi et al., 2015). This variation in frequency suggests differential expression, aggregation or turnover of these peptides.

The significantly lower frequency of poly(AP) and poly(PR) suggests that RAN translation may exhibit a higher efficiency in the sense orientation than in the antisense orientation. The mechanisms underlying this are not yet understood, but one possible explanation could be a differential secondary structure formation by antisense repeats. A low availability of antisense RNA may also limit antisense RAN translation: a lower proportion of neurons have been observed to exhibit antisense RNA foci than sense RNA foci (Gendron et al., 2013; Mizielinska et al., 2013; Zu et al., 2013), and the larger number of antisense foci observed per antisense-foci-positive neuronal nuclei could also reflect a failure to export antisense transcripts from the nucleus following transcription, therefore preventing their translation.

Certain DPR proteins may also be degraded more quickly than others. Indeed, a previous study found that poly(GA) is the only DPR protein that is not degraded by the proteasome (Yamakawa et al., 2014), which would corroborate with the high frequency of poly(GA) inclusions observed. Different DPR proteins may additionally have varying propensities for aggregation. In an immunoassay study performed on *C9orf72* FTD/ALS patient cerebellum, significantly more insoluble poly(GA) was detected than soluble, while the inverse was true for poly(GP) (Gendron et al., 2015). Another study examining protein-only constructs in a

transiently transfected cell model also found that when each DPR protein was individually expressed in cells, poly(GA) was the only peptide that formed the characteristic star-like inclusions, however co-expression of $(GA)_{80}$ and $(GR)_{80}$ resulted in the recruitment of $(GR)_{80}$ into the inclusions (Yang et al., 2015). This suggests that the high frequency of poly(GA) inclusions may be driven by its high propensity for aggregation, with other DPR proteins then recruited into these aggregates to a lesser extent.

It is also interesting that poly(GA) inclusions were observed to be more frequent on average in the cerebellum molecular layer than p62 inclusions. It is possible that this is partially driven by the darker poly(GA) staining, which despite the adjustment of the protocol to account for the differing staining profile, facilitates the detection of additional faint inclusions in comparison to the lighter p62 staining. However, while p62 is thought be present in most inclusions (Mann et al., 2013; Mori et al., 2013c), infrequent p62 negative DPR protein inclusions have been detected in *C9orf72* FTD/ALS patient brain (Schludi et al., 2015), so further verification using double immunostaining would be beneficial to determine whether a substantial subset of poly(GA) inclusions are indeed p62 negative within this region.

The homozygous *C9orf72* FTD case (#20) included in our cohort also exhibited some interesting trends regarding relative inclusion frequencies that may reflect the impact of repeat dosage on DPR protein burden. In both the molecular layer and the granule cell layer, patient #20 exhibited considerably more frequent p62 and poly(GA) inclusions than any other patient, with these inclusions formed in more than double the percentage of cells compared to the mean burden in heterozygous *C9orf72* FTLD patients. However, the same was not observed for poly(GP), poly(GR), poly(AP), or poly(PR) aggregates, which were formed in a more similar proportion of cells in the homozygous case compared to heterozygous *C9orf72* patients on average. This suggests that increased repeat burden

favours poly(GA) inclusion formation, but not necessarily formation of other DPR protein aggregates. In order to study this further it would be interesting to correlate DPR protein inclusion burden with patient repeat length, to determine whether a similar trend is observed.

4.3.7 Caveats for determining relative DPR protein inclusion frequency using immunohistochemistry

When evaluating the different DPR protein inclusion frequencies observed in this immunohistochemical staining, it should also be considered that different antibodies may have varying sensitivities and specificities for their target antigen, which could have a large impact on the number of aggregates detected in patient brain sections. Different antibodies raised against the same protein will also bind different epitopes, which will determine which conformations of the protein are recognised by the antibody. Indeed, in the cell culture work presented in this chapter an anti-poly(GA) antibody recognised insoluble poly(GA) aggregates in the $(GA)_{36}$ and $(GA)_{100}$ transfected HEK293T cells, but not soluble, fluorescently tagged $(GA)_6$-GFP, suggesting that the epitope recognised by this antibody might only be available in the aggregated form of poly(GA). Which conformations the different DPR proteins take in patient brain, how prevalent these conformations are, and which forms are bound by the antibodies used will consequently effect the frequency of DPR inclusions observed in immunostained sections. It is therefore possible that our anti-poly(GA) antibody has a higher sensitivity than our antibodies for other DPR proteins, or that the epitope it recognises is very commonly found in *C9orf72* FTLD patient brain. It may be the case that development of better antibodies for the more infrequently detected DPR protein inclusions, such as poly(PR) and poly(AP), will reveal that these species are more abundant than we currently presume. Other studies have observed similar disparities in abundance among the different DPR proteins, with frequent poly(GA) inclusions, fewer

poly(GP) and poly(GR) inclusions and rare poly(PR) and poly(AP) inclusions (Mackenzie et al., 2015; Schludi et al., 2015). However, effective poly(GA) antibodies may be inherently easier to manufacture than antibodies for poly(PR) or poly(AP), which could lead to the similarity in DPR protein inclusion frequencies observed between groups, and might present one reason why poly(GA) inclusions have thus far been most frequently studied DPR inclusion type within the field. In order to verify the frequencies of DPR proteins observed in this thesis and other studies, future research could be carried out using techniques that do not rely on antibody based detection of proteins, such as mass spectrometry.

4.3.8 DPR protein burden in *C9orf72* FTD/ALS patient cerebellum

We chose to first examine the cerebellum to facilitate the initial optimisation of the analysis protocol, and observed some interesting patterns and differences between the two layers examined that require further thought and investigation. A trend for higher percentage of poly(GA), poly(GP), poly(GR) and poly(AP) inclusion positive cells was observed in the molecular layer compared to the granule cell layer. This was particularly apparent for poly(GR) inclusions, which were present at a more than ten-fold higher frequency in the molecular layer than in the granule cell layer. One factor to be considered regarding this difference is the slight increase in the threshold in DAB staining identification which was required in the granule cell layer due to the presence of dark hematoxylin stain to avoid the false identification of dark portions of nuclei as inclusions. While this could result in a slight reduction of the poly(GR) inclusions identified, an omission of 90 % of all inclusions is unlikely and was not observed when the analysis was checked by eye, arguing that the difference observed is largely genuine. One heterozygous FTD patient (#10) was also observed to exhibit an extremely unequal inclusion burden between the two regions for all DPR proteins and p62: a robust DPR protein inclusion pathology was observed in the cerebellum molecular layer, while inclusions were almost absent from the adjacent granule

cell layer. As these two areas were stained on the same sections at the same time, this limits the likelihood of this difference being due to experimental error. It is not yet clear how these differences in inclusion burden between two regions of the cerebellum might impact a patient's pathology, or what factors might be important to render one region susceptible to DPR protein pathology but an adjacent region invulnerable. As cerebellar atrophy is thought to be a distinguishing feature of *C9orf72* FTD/ALS in comparison to other forms of these diseases(Sha et al., 2012; Whitwell et al., 2012), it would be interesting to investigate how frequent patients with highly mismatched cerebellum DPR burden such as patient #10 are in the *C9orf72* patient population, and to examine whether the presence or absence of inclusions within a sub region of the cerebellum has any impact on patient disease presentation.

Another heterozygous *C9orf72* FTD patient (#13) exhibited unusually few poly(GP) inclusions in either the cerebellum molecular layer or granule cell layer. As this low staining affects both regions on a single stained slide, it would be beneficial to repeat this immunostaining to rule out any experimental error. However, one immunoassay study detected a lower poly(GP) burden specifically in *C9orf72* ALS patient cerebellum (Gendron et al., 2015). While patient #13 was classed pathologically as FTLD, they were previously clinically identified as MND. However, although the single ALS patient included in this cohort exhibited below average poly(GP) burden in both areas, it was not markedly lower than other FTD patients tested. This data set is currently too small to draw any firm conclusions from this trend, and application of this optimised protocol to a larger data set with more patients of differing disease subtypes would therefore be needed to investigate this further.

4.3.9 Further development of the automated DPR protein analysis protocol

Following the examination DPR protein frequency in the cerebellum, this analysis program can now be edited and expanded to function in other brain regions of interest such as the frontal cortex and hippocampus. Additional functionality will be required in order to successfully quantify inclusions within the frontal cortex. Firstly, the greater variability in cell size and hematoxylin intensity within this region will require further optimisation of the cell nuclei identification process. Secondly, additional steps will be required to filter out the more extensive debris and tissue shrinkage artefacts present in this region. Further adaptation to allow distinction between cytoplasmic inclusions, nuclear inclusions and dystrophic neurites would also be desirable to determine the frequency of these entities, especially in the light of a recent study, which found poly(GA) positive dystrophic neurites to correlate with degeneration in patient frontal cortex (Mackenzie et al., 2015).

This protocol could then be applied to a larger cohort of patients across multiple regions in order to examine how these aggregates may affect disease course. In the literature, there have been some conflicting reports as to which inclusions in which areas might correlate with disease subtype. One study found that poly(GA) inclusions were more abundant in the cerebellum granule cell layer in *C9orf72* FTLD cases than in ALS or mixed FTD/MND cases, (Schludi et al., 2015), however this was not replicated by a second study (Mackenzie et al., 2015). Poly(GP) levels were also observed to be lower in the cerebellum in ALS cases compared to FTD or FTD/ALS cases (Gendron et al., 2015), but again this was not found to be the case in another study (Mackenzie et al., 2015). The possibility of increased patient sample size afforded by automated analysis may therefore provide clarity to the question of how DPR protein frequency in different regions may impact disease type.

Chapter 5: The role of the nucleolus in *C9orf72* FTD

5.1 Introduction

The nucleolus is a sub-compartment within the cell nucleus whose principal function is the manufacture, processing and assembly of ribosomal subunits. Nucleoli form around tandemly repeated ribosomal genes known as nucleolar organising regions (NORs), and are comprised of three main sub regions: fibrillar centres, dense fibrillar components and granular components (Boisvert et al., 2007). The fibrillar centres are enriched with a specialised ribosomal RNA (rRNA) polymerase, RNA polymerase I, and transcription of the rDNA repeats occurs primarily at the border between the fibrillar centres and the dense fibrillar components (Boisvert et al., 2007). These transcripts are then modified by small nucleolar ribonuclear proteins (snoRNPs) in the dense fibrillar component, and are finally assembled within the granular component before being exported to the cytoplasm (Boisvert et al., 2007). Due to the involvement of the nucleolus in ribosomal production, it has a critical role to play in the cellular stress response, in which the downregulation of the highly energy-consuming process of rRNA production can be used to reduce cell energy expenditure and limit protein synthesis (Grummt, 2013). However, prolonged inhibition of ribosomal production can result in cellular damage, and nucleolar stress is known to be involved in the stabilisation of p53 under conditions of stress, promoting cell death (Olson, 2004). Consequently, nucleolar dysfunction has been reported in a number of different diseases, including several neurodegenerative disorders (Lee et al., 2014; Parlato and Liss, 2014; Tsoi and Chan, 2014).

Recently, several studies have reported aberrant nucleolar morphology and function in *C9orf72* cell culture models and patient derived cell lines. A number of these models have exhibited a change in nucleolar volume, which is known to reflect a functional change in

rRNA synthesis (Olson, 2011). In particular, the arginine containing DPR proteins poly(GR) and poly(PR), which were shown to cause severe toxicity in our *Drosophila* model (detailed in section 4.2.3), have been implicated as the cause of nucleolar dysfunction. Synthetic $(GR)_{20}$ and $(PR)_{20}$ peptides applied to U2OS cells and cultured astrocytes were found to be taken up by cells, bind to nucleoli, and promote an increase in small nucleolar and ribosomal proteins, a reduction in mature ribosomal subunits, and increased cell death (Kwon et al., 2014). Additionally, $(GR)_{30}$ and $(PR)_{30}$ peptides co-localise with the nucleolar proteins nucleolin and nucleophosmin when overexpressed in cell lines and cause nucleolar swelling and dispersal of nucleophosmin into the nucleoplasm, accompanied by a reduction in the maturation of ribosomal subunits (Tao et al., 2015). Mass spectrometry analysis also showed that $(GR)_{30}$ protein interacts with a number of nucleolar proteins including ribosomal subunits and nucleophosmin (Tao et al., 2015). Overexpression of $(GR)_{50}$ and $(PR)_{50}$ constructs in rat cortical neurons resulted in their co-localisation with nucleolin and fibrillarin (another nucleolar protein), with increased nucleolar area observed (Wen et al., 2014). Furthermore, B-lymphoblasts, fibroblasts, and motor neurons differentiated from *C9orf72* ALS patient derived iPSCs were also observed to exhibit dispersal and fractionation of nucleolin, and a reduction in the maturation of the 45s ribosomal subunit was detected in *C9orf72* ALS motor cortex (Haeusler et al., 2014).

However, despite substantial evidence for nucleolar dysfunction in cell model systems and *in vitro* studies, no evidence for aberrant nucleolar morphology has previously been reported in *C9orf72* patient brain. One study examined nucleolar size in a small sample of two control patient and two *C9orf72* FTD/ALS patient frontal cortices by using ferret diameter in single plane images, but found no difference in diameter between control neurons, *C9orf72* poly(GR) inclusion-positive neurons or *C9orf72* poly(GR) inclusion-negative neurons (Schludi et al., 2015). However, the methodology in this approach lends

itself to a great deal of inaccuracy and may not be sufficient to detect a difference in nucleolar volume.

We therefore carried out a more in depth study to conclusively investigate whether a difference in nucleolar morphology exists in *C9orf72* FTD patient neurons. We also examined whether changes in nucleolar morphology are associated with the presence of poly(GR) inclusions to assess the involvement of arginine-containing DPR proteins in this dysfunction. To achieve this, we examined a larger cohort of eight control and eight heterozygous *C9orf72* FTD patients, plus one homozygous *C9orf72* FTD patient, and captured z-stack images to analyse structures in frontal cortex sections in 3D in order to accurately assess changes in nucleolar volume.

5.2 Results

5.2.1 Immunofluorescent staining of human *C9orf72* FTD frontal cortex

To examine the relationship between neuronal nucleolar morphology and poly(GR) aggregate formation in *C9orf72* FTD, immunofluorescent staining was carried out on human post-mortem *C9orf72* FTD patient and non-neurodegenerative control patient frontal cortical sections. The frontal cortex was examined as this brain region is severely affected by degeneration in *C9orf72* FTD (Snowden et al., 2012). In order to assess the potential involvement of arginine-containing DPR proteins in nucleolar dysfunction, we also examined the presence of poly(GR) inclusions, which are considerably more frequent in *C9orf72* patient tissue than poly(PR) inclusions. To maximise the number of full nucleolar and nuclear volumes that could be visualised and measured, 20 µm thick sections of cortex were stained. Sections underwent triple immunofluorescent staining with the following antibodies: 1) anti-poly(GR) DPR protein primary antibody raised in rat, with an anti-rat Alexa Fluor 546 red fluorescent secondary antibody 2) anti-NeuN neuronal marker primary antibody raised in rabbit, with an anti-rabbit Alexa Fluor 633 far-red fluorescent secondary antibody and 3) either an anti-nucleolin or anti-nucleophosmin nucleolar protein primary antibody raised in mouse, with an anti-mouse Alexa Fluor 488 green fluorescent secondary antibody. Additionally, mountant solution containing DAPI nuclear stain was used to detect chromatin in the blue 405 nm channel. To obtain clear visualisation of nucleolar immunostaining in these thick sections, a robust antigen retrieval step was required to recover antigenic sites masked by cross-linking in the tissue fixation process. When sections underwent antigen retrieval by the standard procedure used by our group of boiling in citrate buffer (pH6) in the microwave, poly(GR) inclusions and NeuN signal could be visualised but nucleolar staining was observed to be faint, present in only a minority of

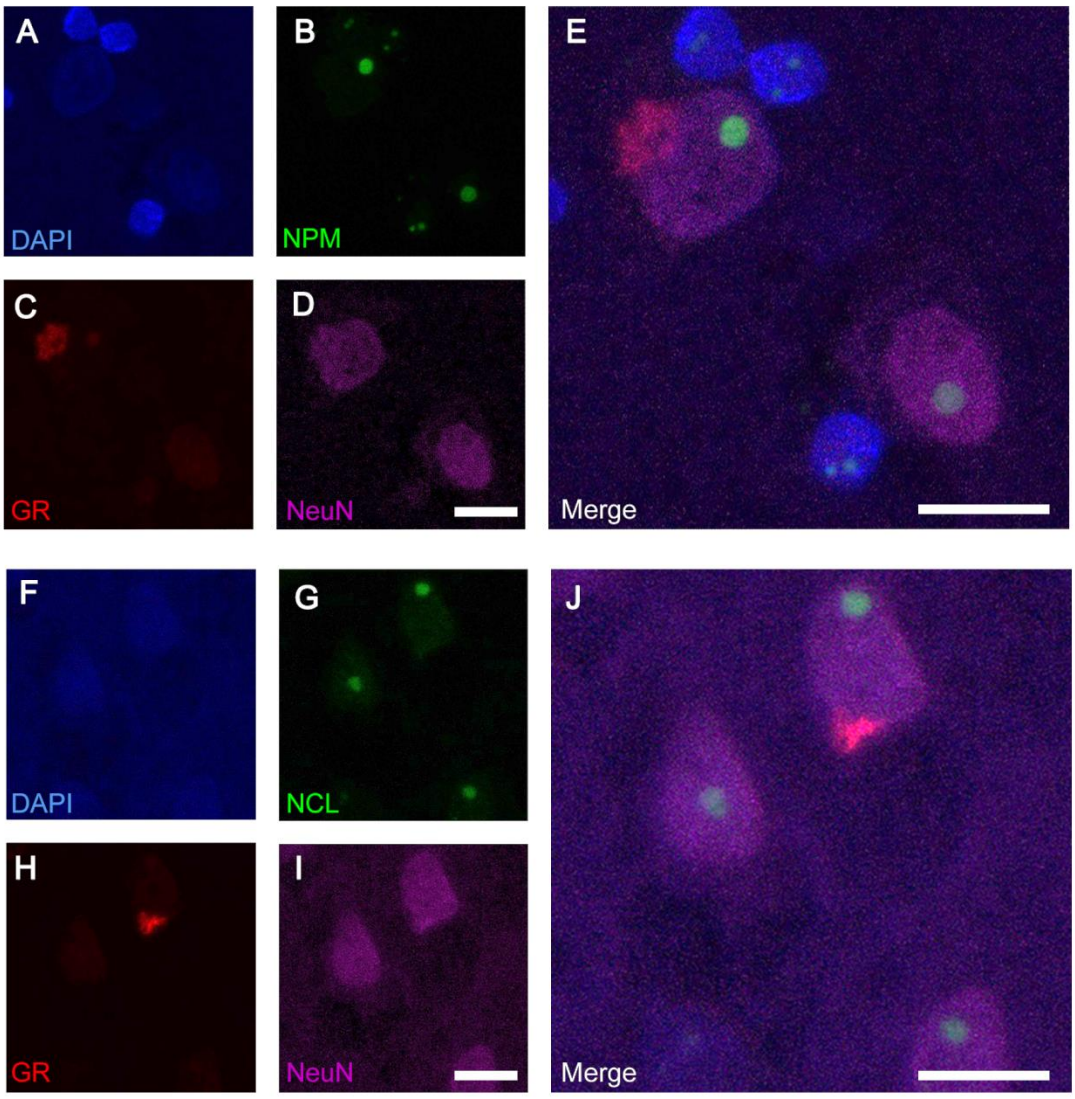


Figure 5.1 Representative images of poly(GR) inclusions and nucleoli in *C9orf72* FTD post mortem brain

Maximum intensity projections of z-stack images from *C9orf72* FTD frontal cortex stained with either (A-E) anti-nucleophosmin (NPM; green) or (F-J) anti-nucleolin (NCL; green) nucleolar marker. All sections were also stained with DAPI nuclear stain (blue), anti-poly(GR) antibody (red), and anti-NeuN neuronal marker (magenta). Scale bars both represent 10 µm in small single channel images or larger merge images. Optimised staining protocol shows clear nucleoli within cell nuclei. Star-like cytoplasmic poly(GR) inclusions are also visible in some neurons.

cells, and variable between different patients. Several different antigen retrieval protocols were therefore tested, and a combined protocol of proteinase K digestion followed by boiling in citrate buffer (pH6) at high pressure within a pressure cooker was found that yielded clear and consistent staining for all antibodies, without excessive damage to tissue architecture. To ensure clear epitope visualisation with each of the primary antibodies used, a range of different primary antibody concentrations were tested, and a

concentration that yielded strong immunofluorescent staining of the target protein with minimal background staining was selected. To develop an image capture protocol for confocal microscopy that could clearly detect immunofluorescent signal in all four fluorescent channels with minimal bleed-through signal from other wavelengths, patient brain sections were singly stained for DAPI and each of the three primary antibodies with their respective secondary antibodies. The range of wavelengths collected for each channel was adjusted such that clear signal was observed from the singly stained section positive for the fluorophore to be detected in that channel, but so that minimal signal was detected for the sections stained with other fluorophores.

5.2.2 Design of automated image analysis protocol

Following the optimisation of the immunofluorescent staining and detection, frontal cortical sections from eight different non-neurodegenerative disease control patients, eight different heterozygous *C9orf72* FTD patients and one homozygous *C9orf72* FTD patient were stained to detect nuclephosmin, poly(GR) inclusions, NeuN and DAPI (Figure 5.1a), with additional sections from the same patients then stained to detect nucleolin, poly(GR) inclusions, NeuN and DAPI (Figure 5.1b). To measure nucleolar volume in neurons with and without poly(GR) inclusions, twenty z-stack images were acquired in each stained section by confocal fluorescence microscopy using a 40 x objective, and an automated image analysis protocol was developed using Volocity image analysis software. DAPI stained nuclei were identified as 3D volumes in the blue 405 nm channel, nucleolar volumes (represented by either nucleolin or nucleophosmin staining) were identified in the green 488 nm channel, poly(GR) inclusion volumes were identified in the red 546 nm channel, and NeuN-positive neuronal volumes were identified in the far-red 633 nm channel (Figure 5.2). Due to slight patient-to-patient variation in staining, DAPI, nucleoli and NeuN populations were identified by measuring signal a set number of standard deviations above the mean voxel intensity for each image. Variation in staining between patients was not

observed to correlate with presence or absence of the *C9orf72* repeat expansion, but may be due to slight differences in tissue preservation and post-mortem delay. As poly(GR) signal is present in *C9orf72* patient sections but not in controls, an absolute intensity value was used to avoid identification of false poly(GR)-positive areas in patients with no inclusions. Different thresholds were used to identify nucleophosmin and nucleolin staining due to slight differences in signal strength, and a higher abundance of low-intensity diffuse nucleolin signal outside of a more defined high-signal-intensity nucleolar compartment. To assess dispersal of nucleolin from the nucleolar compartment into the nucleoplasm, this diffuse nucleolin signal was measured in the green 488 nm channel using a second lower threshold intensity (Figure 5.3). The nucleolar population was then subtracted from the diffuse nucleolin population to give the volume of nucleolin within the nucleoplasm.

Due to a high level of background fluorescence observed inside nucleoli in the red 546 nm channel in both control patient and *C9orf72* FTD patient sections, the nucleolar population was subtracted from the poly(GR) population, such that any poly(GR) signal present solely within a nucleolus was not counted. No genuine poly(GR) inclusions contained solely within

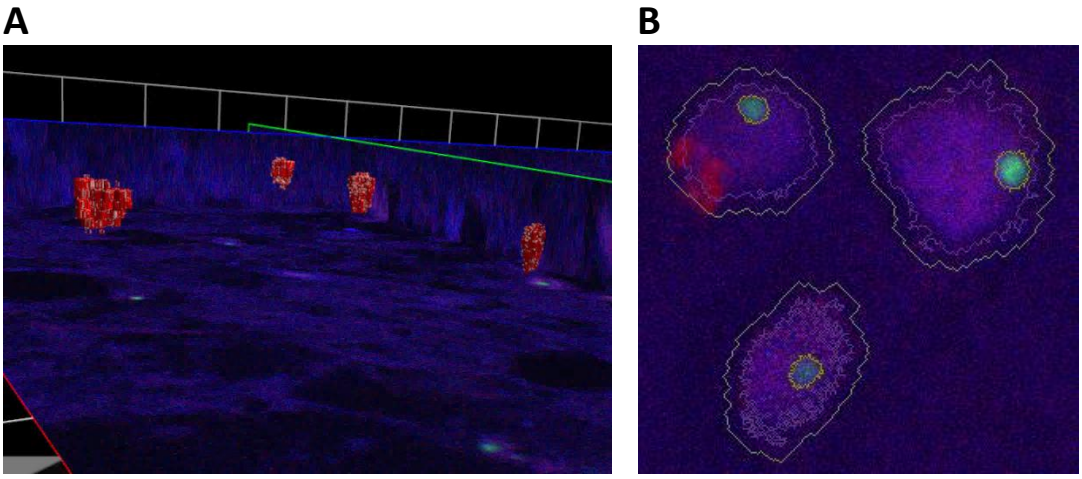


Figure 5.2 Volocity volumetric image analysis

A) 3D representation of GR inclusion objects (red) identified within a z-stack image by Volocity image analysis software. B) Maximum intensity projection of nucleophosmin stained neurons showing identification of objects as nuclei (purple line), nucleoli (yellow line), GR inclusions (red line) and NeuN positive neurons (green line).

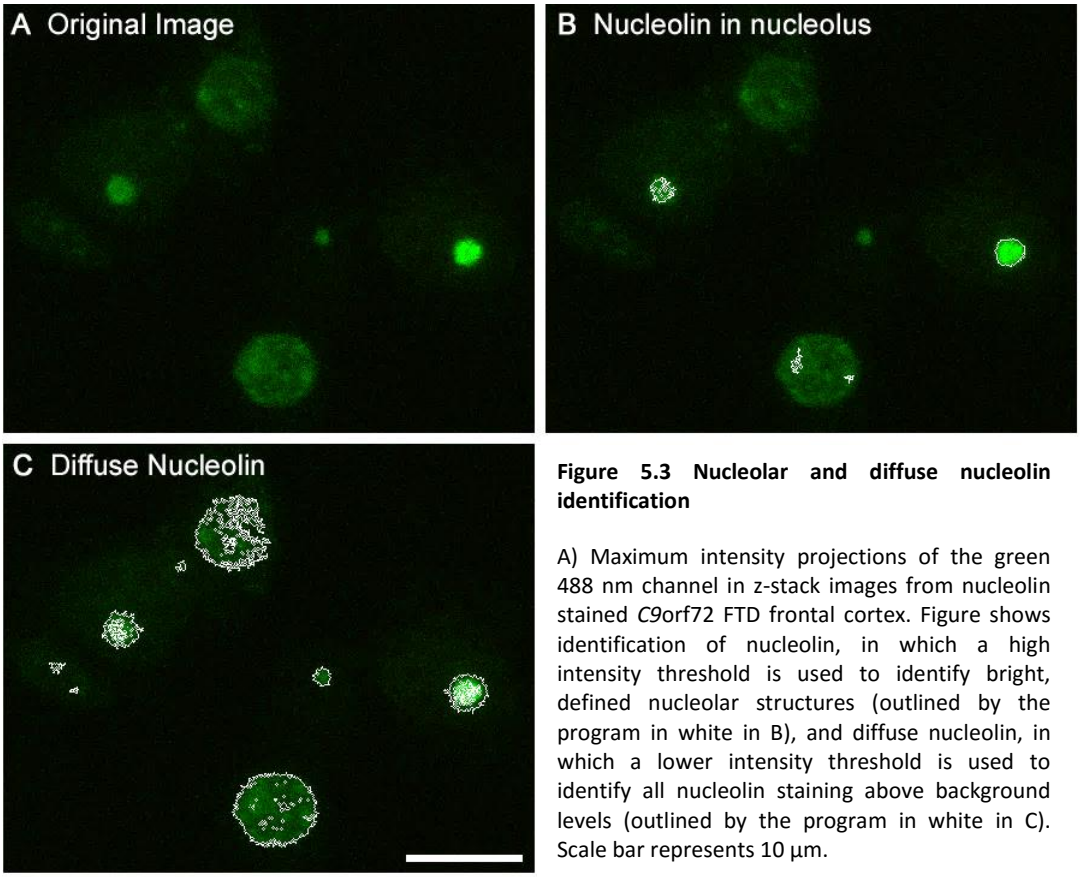


Figure 5.3 Nucleolar and diffuse nucleolin identification

A) Maximum intensity projections of the green 488 nm channel in z-stack images from nucleolin stained *C9*orf72 FTD frontal cortex. Figure shows identification of nucleolin, in which a high intensity threshold is used to identify bright, defined nucleolar structures (outlined by the program in white in B), and diffuse nucleolin, in which a lower intensity threshold is used to identify all nucleolin staining above background levels (outlined by the program in white in C). Scale bar represents 10 µm.

the nucleolus were evident above background staining levels upon examination of the images by eye.

To identify neuronal nuclei, a new population was made containing DAPI-stained nuclei that were also positive for NeuN staining. These areas were then dilated by 10 pixels to encompass cytoplasmic inclusions not touching the nucleus, and DAPI, nucleolar and poly(GR) volumes touching these areas were compartmentalised as belonging to a specific neuron. DAPI, nucleolar or poly(GR) volumes that touched more than one neuron were assigned to the neuron to which their centroid was closest. The analysis protocol yielded a list of neurons present in each image, with the total volume of nucleolar staining, DAPI staining and poly(GR) staining compartmentalised in each neuron. Tables of neurons were compiled for each patient: any neurons that did not contain any DAPI signal or nucleolar signal were removed from the analysis, and neurons were then divided into two groups

	Control patients	1	2	3	4	5	6	7	8
Nucleophosmin	No. GR- cells	520	585	871	456	561	605	576	882
	No. false positives	1	4	4	1	1	2	0	2
	Total no. cells	521	589	875	457	562	607	576	884
	% false positives	0.19%	0.68%	0.46%	0.22%	0.18%	0.33%	0.00%	0.23%
Nucleolin	No. GR- cells	513	520	669	344	265	261	232	634
	No. false positives	0	0	0	0	0	0	0	0
	Total no. cells	513	520	669	344	265	261	232	634
	% false positives	0%	0%	0%	0%	0%	0%	0%	0%

Table 5.1 Details of neurons analysed in control patient cortices

GR- = poly(GR) inclusion-negative

	C9 FTD Patients	10	11	12	14	16	17	18	19	20
Nucleophosmin	No. GR- cells	964	427	784	720	696	836	434	721	634
	No. GR+ cells	26	23	101	18	53	50	39	20	139
	Total no. cells	990	450	885	738	749	886	473	741	773
	% GR+ cells	2.63%	5.11%	11.41%	2.44%	7.08%	5.64%	8.25%	2.70%	17.98%
Nucleolin	No. GR- cells	616	469	519	584	688	503	198	525	378
	No. GR+ cells	17	24	66	16	61	26	19	21	92
	Total no. cells	633	493	585	600	749	529	217	546	470
	% GR+ cells	2.69%	4.87%	11.28%	2.67%	8.14%	4.91%	8.76%	3.85%	19.57%
	TOTAL GR+ CELLS	**2.65%**	**4.98%**	**11.36%**	**2.54%**	**7.61%**	**5.37%**	**8.41%**	**3.19%**	**18.58%**

Table 5.2 Details of neurons analysed in *C9*orf72 FTD patient cortices

GR- = poly(GR) inclusion-negative, GR+ = poly(GR) inclusion-positive

that were either positive or negative for poly(GR) signal. *C9orf72* patient sections contained between 16 and 139 poly(GR)-positive neurons within the images taken (Table 5.2).

In heterozygous *C9orf72* FTD patients poly(GR) inclusions were present in an average of 5.8 ± 1.1 % of neurons (mean ± SEM). A notably high poly(GR) inclusion burden was observed in the single homozygous *C9orf72* FTD patient (#20), with 18.6 % of neurons exhibiting a poly(GR) inclusion in the frontal cortex, which could be ascribed to the additional expanded repeat allele present in this patient.

The majority of control patients were completely absent for poly(GR) positive neurons, however between 1 and 4 positive neurons were detected in 7 out of 8 control patient sections stained with nucleophosmin (Table 5.1). Out of a total 5071 nucleophosmin stained control neurons analysed, 15 were identified as poly(GR) positive, representing 0.30 % of the cell population. When these images were examined by eye, it was confirmed that these were false positives, which often occurred due to the presence of residual autofluorescent debris on the slide despite the use of a sudan black autofluorescence quenching step. A similar false positive error rate may also be expected in the poly(GR) positive neuron population in *C9orf72* patients, but would represent as little as 1 to 5 % of the total inclusion burden. Out of a total 3438 nucleolin stained control neurons analysed, none were identified as poly(GR) positive.

5.2.3 Nucleolar volume

To examine the distribution of nucleolar sizes within the neuronal populations analysed, frequency histograms were plotted for the nucleophosmin and nucleolin volumes present in the neuronal populations from each *C9orf72* FTD patient brain (Figure 5.4). Additionally, data for all control patient neurons, *C9orf72* FTD patient poly(GR)-inclusion-positive neurons and *C9orf72* FTD poly(GR)-inclusion-negative neurons were pooled and plotted (Figure 5.5). A non-normal, positively-skewed distribution was observed for nucleolar volume. The majority of individual patient histograms show a positive shift in both nucleophosmin and nucleolin volume in poly(GR)-inclusion-positive neurons compared to poly(GR)-inclusion-negative neurons, suggesting that poly(GR) inclusions are associated with an increased nucleolar volume (Figure 5.4). When data for all patients was pooled, there was also a slight shift to the right for nucleophosmin and nucleolin volume in neurons that contained poly(GR) inclusions compared to poly(GR)-negative neurons in both *C9*orf72 FTD patients and control patients (Figure 5.5).

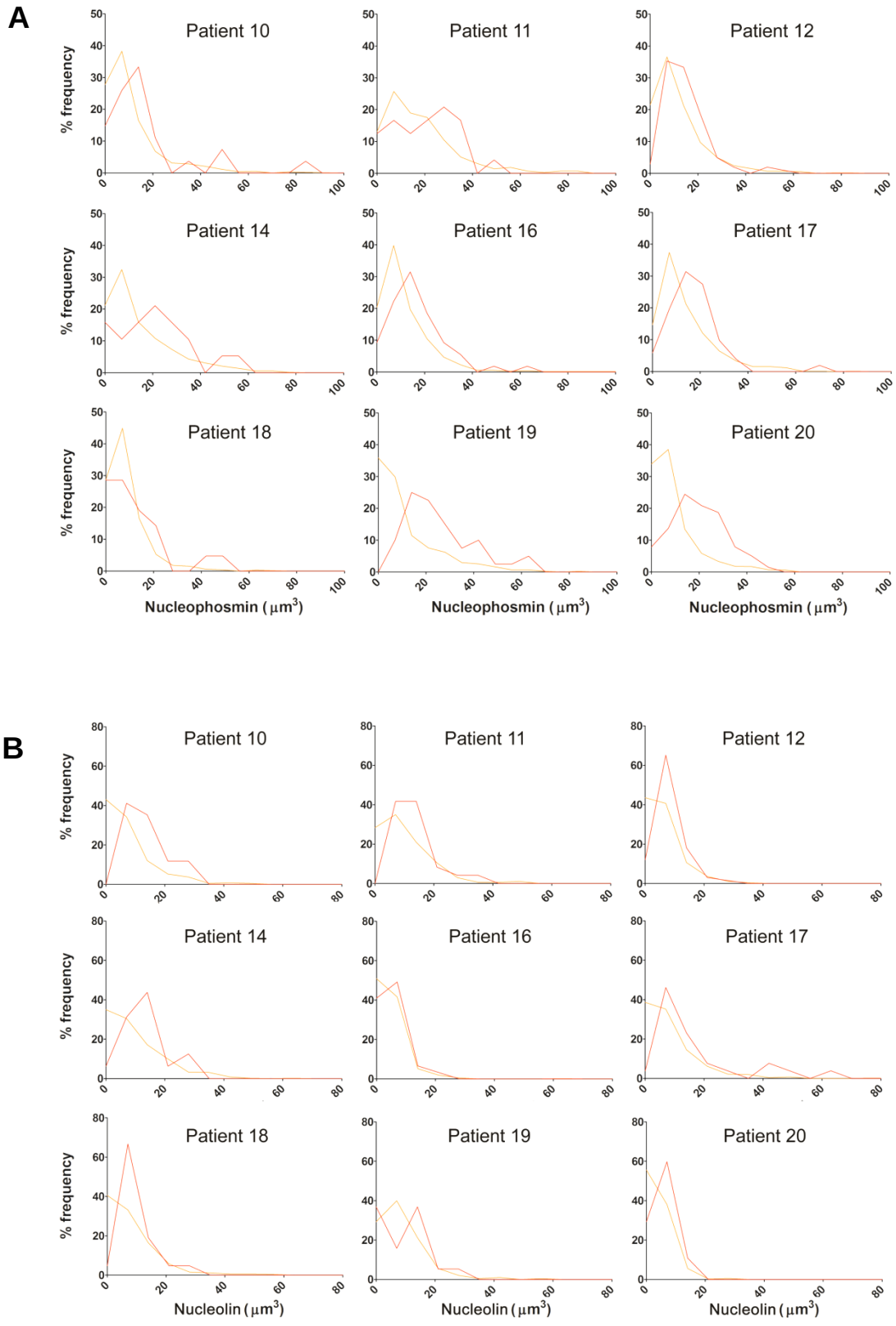


Figure 5.4 Frequency distribution of nucleolar volumes in individual post-mortem patient cortex

Frequency histograms of nucleolar volumes from poly(GR)-inclusion-positive (red) and poly(GR)-inclusion-negative (orange) neurons, in either A) nucleoophosmin stained or B) nucleolin stained sections from each individual *C9orf72* FTD patient.

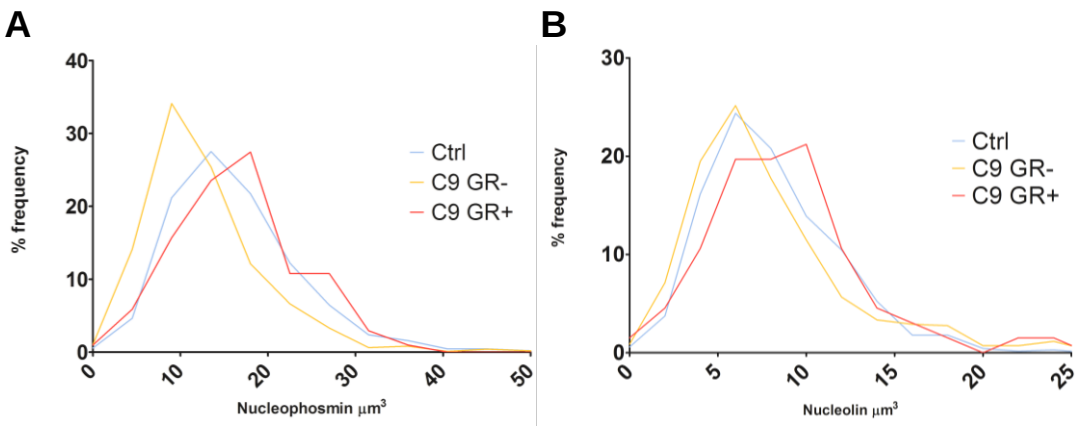


Figure 5.5 Pooled frequency distribution of nucleolar volumes in post-mortem patient cortex

Histograms displaying nucleolar volumes pooled from all control patients (blue), GR-inclusion-negative neurons (C9 GR-) from all *C9orf72* FTD patients (orange) and GR-inclusion-positive (C9 GR+) neurons from all *C9orf72* FTD patients (red) for either A) nucleoophosmin stained or B) nucleolin stained sections from each individual *C9orf72* FTD patient.

We next performed two statistical tests to answer the following questions: 1) Is there a significant difference in nucleolar volume between control patient neurons, and poly(GR)-inclusion-negative neurons from *C9orf72* FTD patients? 2) Is there a significant difference in nucleolar volume between poly(GR)-inclusion-positive neurons and poly(GR)-inclusion-negative neurons in *C9orf72* FTD patients? Due to the non-normal distribution of the nucleolar volumes, median nucleolar volume was plotted as a location parameter for each patient (Figure 5.6).

To test whether there is a significant difference in median nucleolar volume in poly(GR)-inclusion-negative neurons from *C9orf72* FTD patients compared to control neurons, an unpaired students t-test was carried out on these two groups. While no significant difference was observed between nucleolin volume in control and *C9orf72* FTD poly(GR)-negative cells ($p = 0.284$), a significant decrease was observed in nucleophosmin volume in *C9orf72* FTD poly(GR)-inclusion-negative neurons compared to control patient neurons ($p < 0.05$).

Next we looked to see if there was a statistically significant difference in nucleolar volume in poly(GR)-inclusion-positive neurons in *C9orf72* FTD patients compared to poly(GR)-

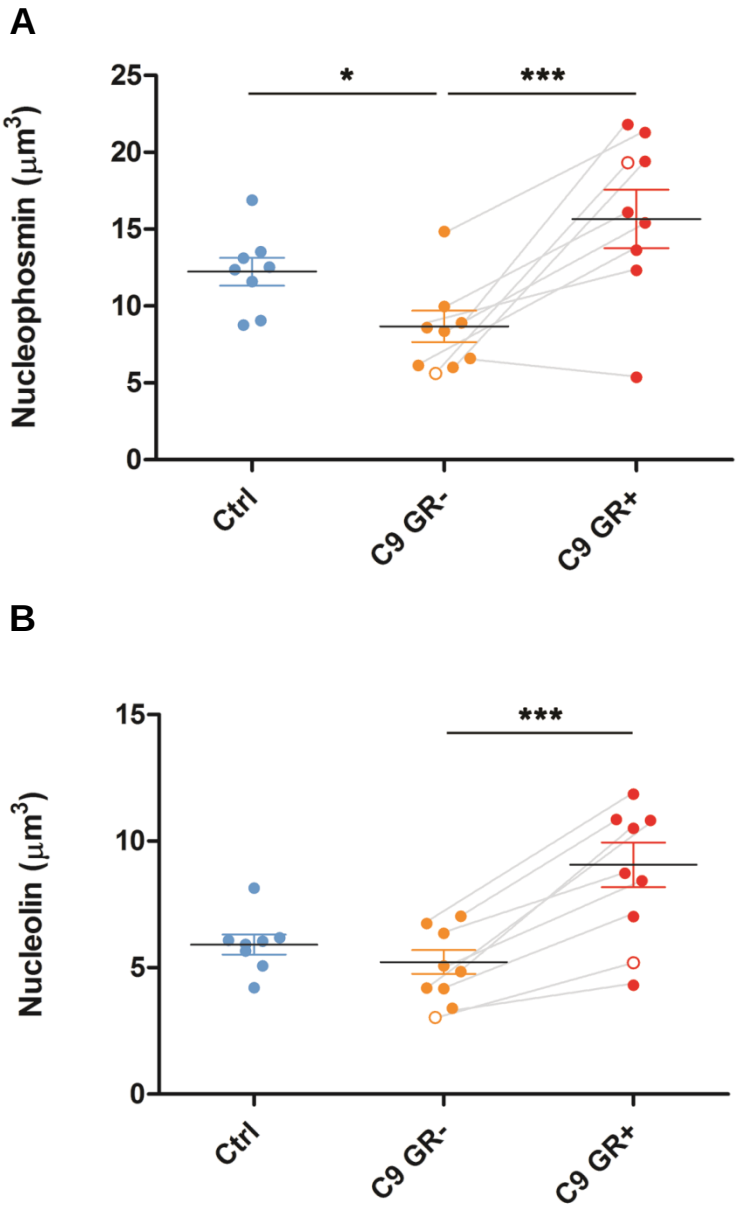


Figure 5.6 Median neuronal nucleolar volumes in post-mortem frontal cortex

Quantification of neuronal nucleolar volume as determined by (a) nucleophosmin or (b) nucleolin immunoreactivity. Three categories of neurons were analysed - control patient neurons (Ctrl), *C9orf72* FTD patient neurons without a poly(GR) inclusion (C9 GR-), and *C9orf72* FTD patient neurons with a poly(GR) inclusion (C9 GR+). Each point represents the median nucleolar volume for a single patient. Grey lines show pairing between the same *C9orf72* FTD individuals in poly(GR)-inclusion-positive and poly(GR)-inclusion-negative groups. Empty circle represents homozygous *C9orf72* FTD patient, not included in mean calculation or statistical analyses. Black lines show the mean of patient medians for each group. Error bars represent standard error of mean (SEM). *p<0.05, ***p<0.0001

inclusion-negative neurons in *C9orf72* FTD patients. To do this we carried out a paired regression analysis on these two groups. Use of a regression analysis incorporating all individual nucleolar measurements accounts for the unequal variability and error associated with the poly(GR)-inclusion-positive and inclusion-negative median nucleolar volumes in *C9orf72* FTD patients, due to the fact that poly(GR) inclusions are only present in a minority of cells. Additionally, paired analysis accounts for the fact that the poly(GR)-inclusion-positive and poly(GR)-inclusion-negative neurons are derived from the same *C9orf72* FTD patients, and a paired relationship therefore exists between nucleolar volumes derived from the same patient. This relationship is apparent when we examine the difference in median nucleolar volume for each individual *C9orf72* FTD patient: for all patients, median nucleophosmin volume is higher in poly(GR)-inclusion-positive than inclusion-negative neurons, while median nucleolin volume is higher in inclusion-positive than inclusion-negative neurons in all patients except one. A statistically significant increase was observed in both nucleophosmin and nucleolin volume in *C9orf72* FTD patient neurons with poly(GR) inclusions compared to those without poly(GR) inclusions (mean ± SEM: nucleophosmin: 15.5 ± 2.0 μm^3 vs 8.7 ± 1.0 μm^3, $p < 0.0001$; nucleolin: 9.0 ± 0.9 μm^3 vs 5.2 ± 0.5 μm^3; $p < 0.0001$).

5.2.4 Nucleolar size relative to nuclear size

As a lower median nucleophosmin volume was observed in *C9orf72* FTD patient poly(GR)-inclusion-negative cells compared to control cells, this raises the possibility that poly(GR) inclusions are present within a subset of the neuronal population with larger nucleoli. It is also possible that the increase in nucleolar volume observed is part of a more generic increase in cell size. To help determine whether these factors might be driving the observed differences in nucleolar volume, we examined nuclear volume in each of the cell populations. Cells with a larger nucleus might be expected to also have a larger nucleolus. To examine this possibility, we correlated neuronal nuclear volume against nucleophosmin

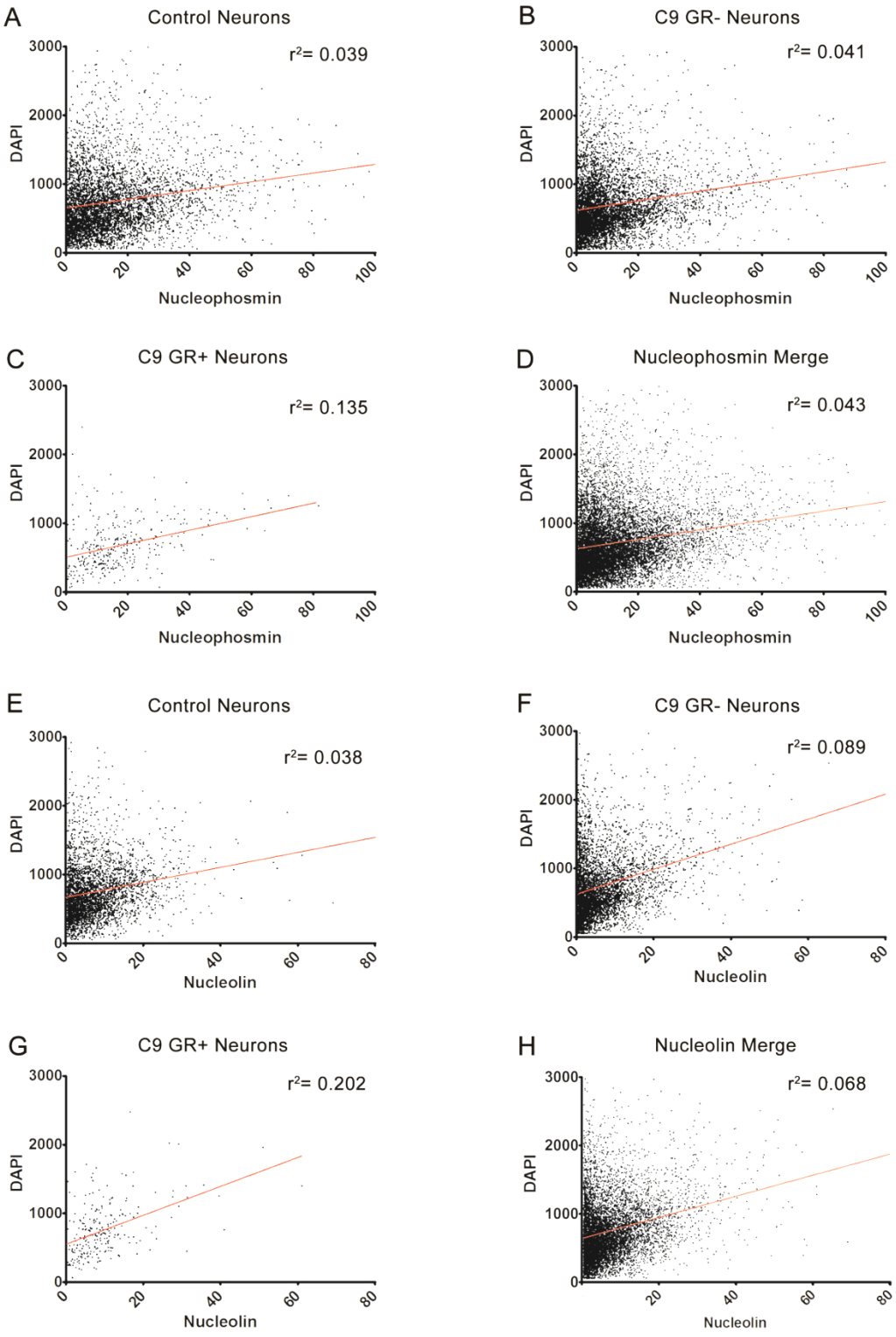


Figure 5.7 Correlation of nucleolar volume with nuclear volume

DAPI nuclear volume was plotted against nucleophosmin volume (A-D) or nucleolin volume (E-H) for each neuron measured. Linear regressions were carried out on control neurons (A,E), *C9orf72* FTD patient poly(GR)-inclusion-negative neurons (C9 GR-; B,F), *C9orf72* FTD patient poly(GR)-inclusion-positive neurons (C9 GR+; C,G), and neurons pooled from all three populations (D,H). Linear regressions were carried out on each comparison. Red lines show line of best fit. All regression lines were found to have a significantly non-zero slope $p < 0.0001$. r^2 values represent the goodness of fit of the data to the regression line.

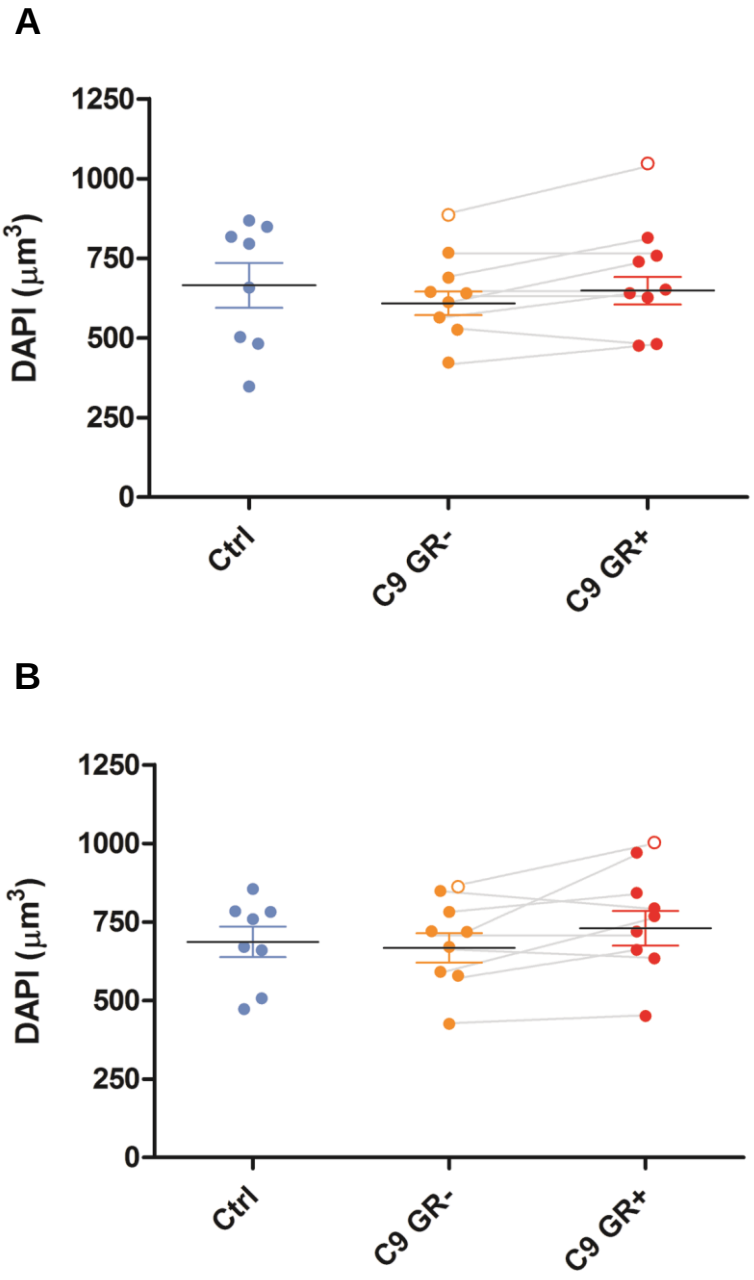


Figure 5.8 Median neuronal nuclear volume in post-mortem frontal cortex

Quantification of neuronal nuclear volume as determined by DAPI staining in (a) nucleophosmin stained slides or (b) nucleolin stained slides. Three categories of neurons were analysed - control patient neurons (Ctrl), *C9orf72* FTD patient poly(GR)-inclusion-negative neurons (C9 GR-), and *C9orf72* FTD patient poly(GR)-inclusion-positive neurons (C9 GR+). Each point represents the median nuclear volume for a single patient. Grey lines show pairing between the same C9FTD individuals in C9 GR- and C9 GR+ groups. Empty circle represents C9 homozygous patient, not included in mean calculation or statistical analyses. Black lines show the mean of patient medians for each group. Error bars represent standard error of mean (SEM).

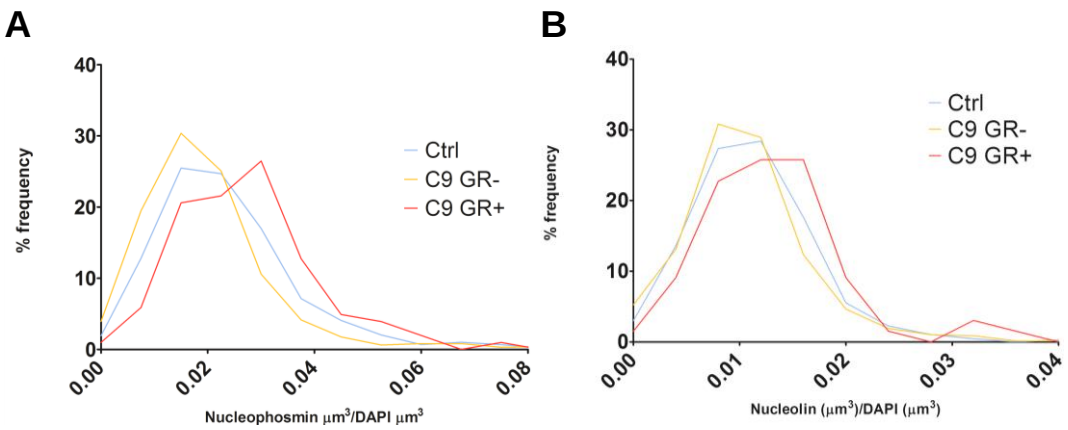


Figure 5.9 Pooled frequency distribution of nucleolar proportion of nucleus in post-mortem patient cortex

Histograms displaying nucleolar volume as a proportion of nuclear volume, pooled from all control patients (ctrl; blue), poly(GR)-inclusion-negative neurons from all *C9orf72* FTD patients (C9 GR-; orange) and poly(GR)-inclusion-positive neurons from all *C9orf72* FTD patients (C9 GR+; red) for either A) nucleoophosmin stained or B) nucleolin stained sections from each individual *C9orf72* FTD patients.

or nucleolin volume (Figure 5.7). Linear regressions revealed that nuclear volume is weakly positively correlated against nucleolar volume in neurons. Lines of best fit plotted for these populations had a significantly non-zero slope ($p < 0.001$) with low r^2 values ranging between 0.038 and 0.202 for the different sub groups signifying weak correlation.

Next, we examined whether there was a difference in median nuclear size between control patient neurons, *C9orf72* FTD patient poly(GR)-inclusion-negative neurons and *C9orf72* FTD patient poly(GR)-inclusion-positive neurons (Figure 5.8). No significant difference in nuclear volume was observed between *C9orf72* FTD patient poly(GR)-inclusion-negative and *C9orf72* FTD patient poly(GR)-inclusion-positive neurons by paired regression analysis for either nucleophosmin ($p = 0.963$) or nucleolin ($p = 0.168$) stained sections. Similarly, no significant difference was observed in nuclear volume between control patient neurons and *C9orf72* FTD patient poly(GR)-inclusion-negative neurons by unpaired t-test for either nucleophosmin ($p = 0.491$) or nucleolin ($p = 0.781$) stained sections.

Finally, we examined whether the nucleolus occupies a larger proportion of total nuclear volume in *C9orf72* FTD patient poly(GR)-inclusion-positive neurons compared to *C9orf72* FTD patient poly(GR)-inclusion-negative neurons and control patient neurons. To do this,

total nucleolar volume was divided by total nuclear volume for each neuron in both nucleophosmin and nucleolin stained sections. Frequency histograms were plotted for the distributions of the nuclear proportion of nucleophosmin and nucleolin in *C9orf72* FTD patient poly(GR)-inclusion-positive, *C9orf72* FTD patient poly(GR)-inclusion-negative and control patient neurons pooled from all patients (Figure 5.9). As was observed for absolute nucleolar volume, *C9orf72* FTD patient poly(GR)-inclusion-positive neurons had a positive shift in both nucleophosmin and nucleolin proportion of the nucleus compared to *C9orf72* FTD patient poly(GR)-inclusion-negative neurons and control neurons.

Next we looked to see if there was a statistically significant increase in nucleolar proportion of the nucleus in *C9orf72* FTD patient poly(GR)-inclusion-positive cells compared to *C9orf72* FTD patient poly(GR)-inclusion-negative cells and median values were plotted for each patient (Figure 5.10). Nucleophosmin occupied a significantly larger proportion of the nucleus in *C9orf72* FTD patient poly(GR)-inclusion-positive neurons compared to *C9orf72* FTD patient poly(GR)-inclusion-negative neurons when assessed by paired regression analysis ($p < 0.01$), however while a trend for a larger nuclear proportion was observed for nucleolin, paired regression analysis showed that this was not significant ($p = 0.122$). No significant difference was observed in nucleolar proportion of the nucleus between *C9orf72* FTD patient poly(GR)-inclusion-negative neurons and control patient neurons as assessed by unpaired t-test for either nucleophosmin ($p = 0.144$) or nucleolin ($p = 0.681$).

Therefore while an increase in nucleolar volume in proportion to nuclear volume is observed in *C9orf72* FTD patient poly(GR)-inclusion-positive neurons compared to *C9orf72* FTD patient poly(GR)-inclusion-negative neurons, this trend is less robust compared to the increase in absolute nuclear volume observed.

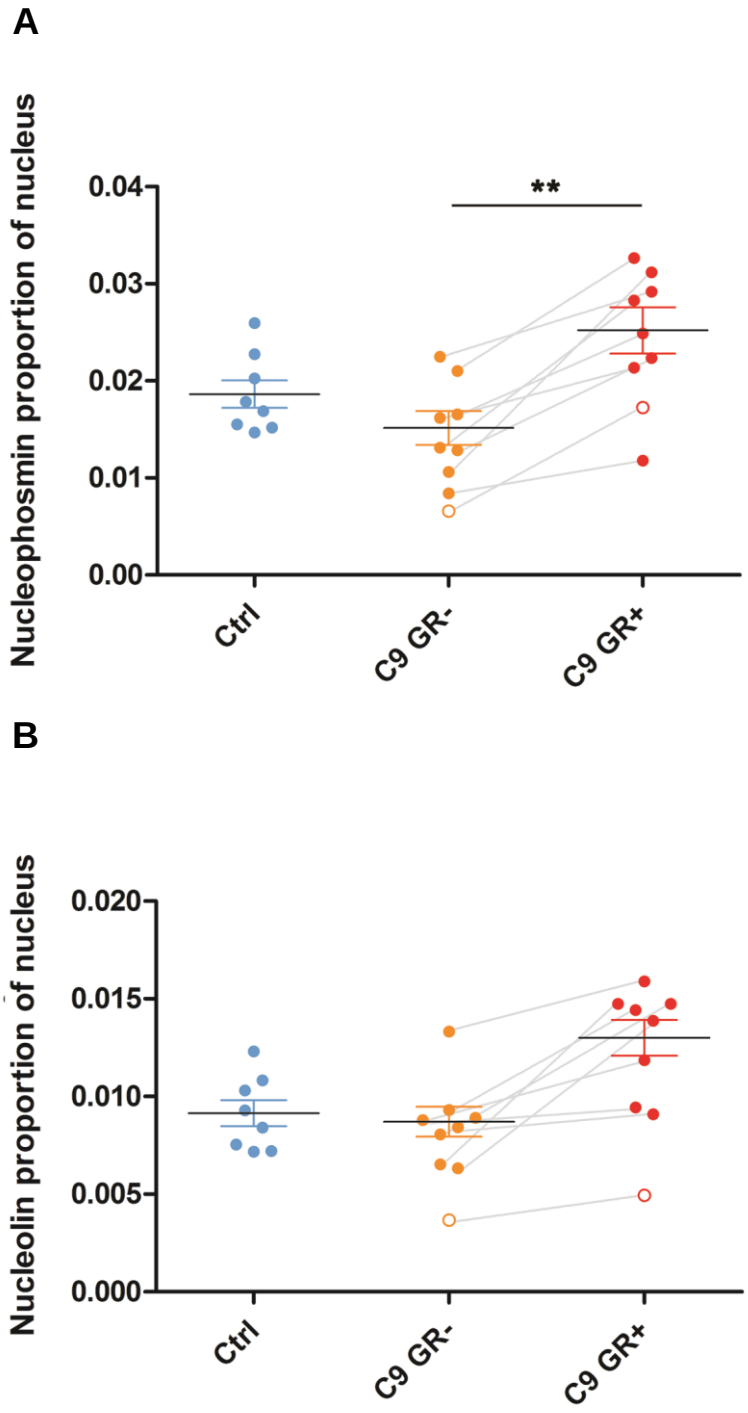


Figure 5.10 Median neuronal nucleolar proportion of nucleus in post-mortem frontal cortex

Quantification of neuronal nucleolar proportion of nucleus as determined by (a) nucleophosmin or (b) nucleolin immunoreactivity. Three categories of neurons were analysed - control patient neurons (Ctrl), *C9orf72* FTD neurons without a poly(GR) inclusion (C9 GR-), and *C9orf72* FTD neurons with a poly(GR) inclusion (C9 GR+). Nucleolar volume was divided by nuclear volume (as determined by DAPI staining) in each neuron. Each point represents the median nucleolar proportion of nucleus for a single patient. Grey lines show pairing between the same *C9orf72* FTD individuals in C9 GR- and C9 GR+ groups. Empty circle represents homozygous *C9orf72* FTD patient, not included in mean calculation or statistical analyses. Black lines show the mean of patient medians for each group. Error bars represent standard error of mean (SEM). **p<0.01

5.2.5 Nucleolar dispersal and fragmentation

We next examined whether *C9orf72* FTD patient neurons exhibit a diffuse or fragmented localisation of nucleolar markers, as has been observed in *C9orf72* cell models and patient lymphoblasts (Haeusler et al., 2014; Tao et al., 2015). When nucleophosmin staining was examined, no signal above background intensity could be observed outside of defined nucleoli (Figure 5.1). Conversely, nucleolin staining was observed to consist of both a defined nucleolar compartment and a fainter, more diffuse signal within the nucleus (Figure 5.3). To determine whether there is a difference in the volume of diffuse nucleolin signal, a lower threshold intensity was used to identify staining in the green 488 channel than was used to measure the defined nucleolar structures. The nucleolar population was then subtracted from the diffuse nucleolin population to leave only low threshold staining. Median volume of diffuse staining per neuron was plotted for control patient neurons, *C9orf72* FTD patient poly(GR)-inclusion-positive neurons and *C9orf72* FTD patient poly(GR)-inclusion-positive (Figure 5.11a). No difference was observed between control patient neurons and *C9orf72* FTD patient poly(GR)-inclusion-negative neuron, or between *C9orf72* FTD patient poly(GR)-inclusion-positive neurons and *C9orf72* FTD patient poly(GR)-inclusion-negative neurons. As the volume of diffuse nucleolin may be associated with the nuclear volume of each neuron, we next examined the proportion of each nucleus occupied by diffuse nucleolar staining, but similarly, no difference was observed between control patient neurons and *C9orf72* FTD patient poly(GR)-inclusion-negative neurons, or between *C9orf72* FTD patient poly(GR)-inclusion-positive neurons and *C9orf72* FTD patient poly(GR)-inclusion-negative neurons (Figure 5.11b).

To investigate whether a nucleolar fragmentation phenotype is exhibited by nucleolin and nucleophosmin, we next examined the number of nucleolar puncta per neuron. The number of distinct nucleophosmin stained or nucleolin stained objects in each neuron was analysed. Control patient neurons, and *C9orf72* FTD patient poly(GR)-inclusion-positive

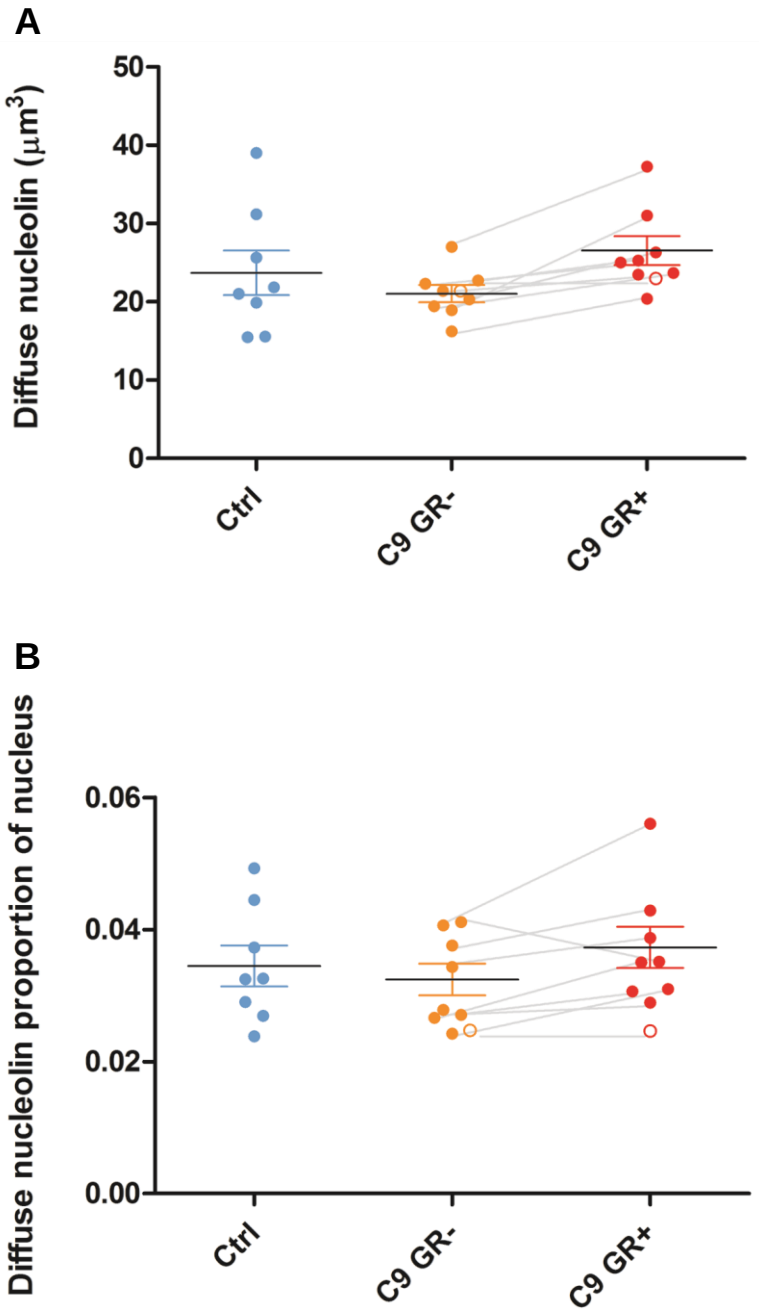


Figure 5.11 Median diffuse nucleolin volumes in post-mortem frontal cortex

Quantification of (a) diffuse nucleolin volume and (b) difuse nucleolin volume as a proportion of DAPI nuclear volume. Three categories of neurons were analysed - control patient neurons (Ctrl), *C9orf72* FTD neurons without a poly(GR) inclusion (C9 GR-), and *C9orf72* FTD neurons with a poly(GR) inclusion (C9 GR+). Each point represents (a) the median diffuse nucleolin volume for a single patient or (b) the median diffuse nucleolin proportion of the nucleus for a single patient. Grey lines show pairing between the same *C9orf72* FTD individuals in poly(GR)-inclusion-positive and poly(GR)-inclusion-negative groups. Empty circle represents homozygous *C9orf72* FTD patient, not included in mean calculation or statistical analyses. Black lines show the mean of patient medians for each group. Error bars represent standard error of mean (SEM).

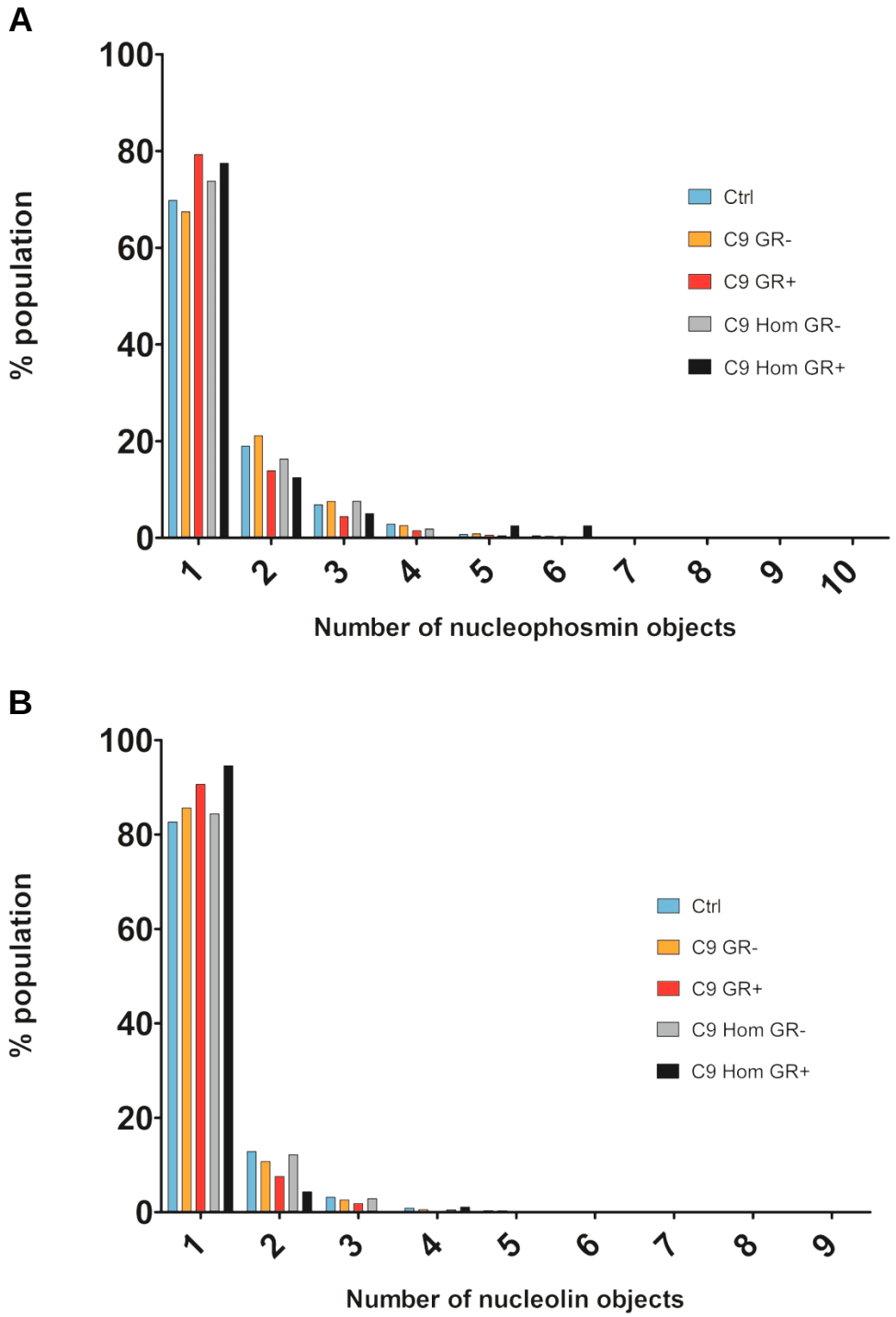


Figure 5.12 Frequency distribution of number of distinct nucleolar objects per neuron in post-mortem patient frontal cortex

Histograms displaying the proportion of the population exhibiting a given number of defined nucleolar objects per cell, pooled from all control patients (Ctrl; blue), poly(GR)-inclusion-negative neurons from all heterozygous *C9orf72* FTD patients (C9 GR-; orange,) poly(GR)-inclusion-positive neurons from all heterozygous *C9orf72* FTD patients (C9 GR+; red) poly(GR)-inclusion-negative neurons for the single homozygous *C9orf72* FTD patient (C9 Hom GR-; grey) , and poly(GR)-inclusion-positive neurons for the single homozygous *C9orf72* FTD patient (C9 Hom GR+; black) for either A) nucleophosmin stained or B) nucleolin stained sections.

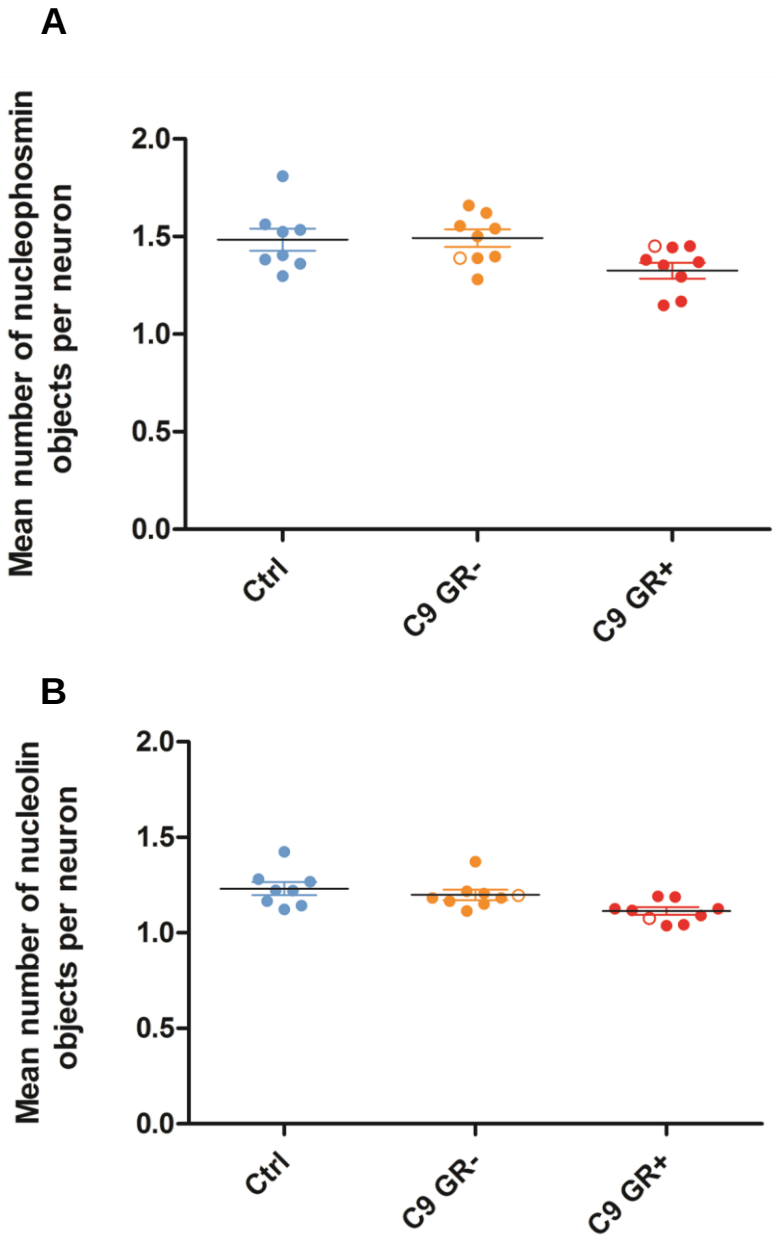


Figure 5.13 Mean number of distinct nucleolar objects per neuron in post-mortem patient frontal cortex

Quantification of distinct nucleolar objects per neuron as determined by (a) nucleophismin or (b) nucleolin immunoreactivity. Three categories of neurons were analysed - control patient neurons (Ctrl), *C9orf72* FTD neurons without a poly(GR) inclusion (C9 GR-), and *C9orf72* FTD neurons with a poly(GR) inclusion (C9 GR+). Each point represents the mean number of objects for a single patient. Empty circle represents C9 homozygous patient, not included in mean calculation or statistical analyses. Black lines show the grand mean of patient means for each group. Error bars represent standard error of mean (SEM).

neurons and *C9orf72* FTD patient poly(GR)-inclusion-negative neurons were pooled from all patients, and a histogram was plotted displaying the proportion of the neuronal population containing a given number of nucleolar structures in nucleophosmin and nucleolin stained slides (Figure 5.12). No difference in the number of nucleoli between control patient neurons, and *C9orf72* FTD patient poly(GR)-inclusion-negative neurons or and *C9orf72* FTD patient poly(GR)-inclusion-positive neurons was observed in cells stained with nucleolin or nucleophosmin, with over 60 % of each population containing one nucleolus. This was further confirmed by plotting the mean number of nucleoli per cell for each patient, with no difference observed between control patient neurons, *C9orf72* FTD patient poly(GR)-inclusion-positive neurons or *C9orf72* FTD patient poly(GR)-inclusion-negative neurons (Figure 5.13).

5.2.6 Nucleolar dysfunction and disease course

Finally, we investigated the potential for a correlation between increased nucleolar volume in *C9orf72* FTD patient poly(GR)-inclusion-positive neurons and age of disease onset or disease duration in each *C9orf72* FTD patient. Two values were used to examine nucleolar volume for both nucleolin and nucleophosmin immunostained nucleoli in each *C9orf72* FTD patient: 1) median nucleolar volume in *C9orf72* FTD patient poly(GR)-inclusion-positive neurons, and 2) the difference in median nucleolar volume between *C9orf72* FTD patient poly(GR)-inclusion-negative and *C9orf72* FTD patient poly(GR)-inclusion-positive neurons. These measurements of nucleolar volume were each plotted against age of disease onset and disease duration of each patient to give eight total comparisons (Figure 5.1.4). A linear regression analysis was carried out on each pair of factors, however, no significant correlation was observed for any of the comparisons.

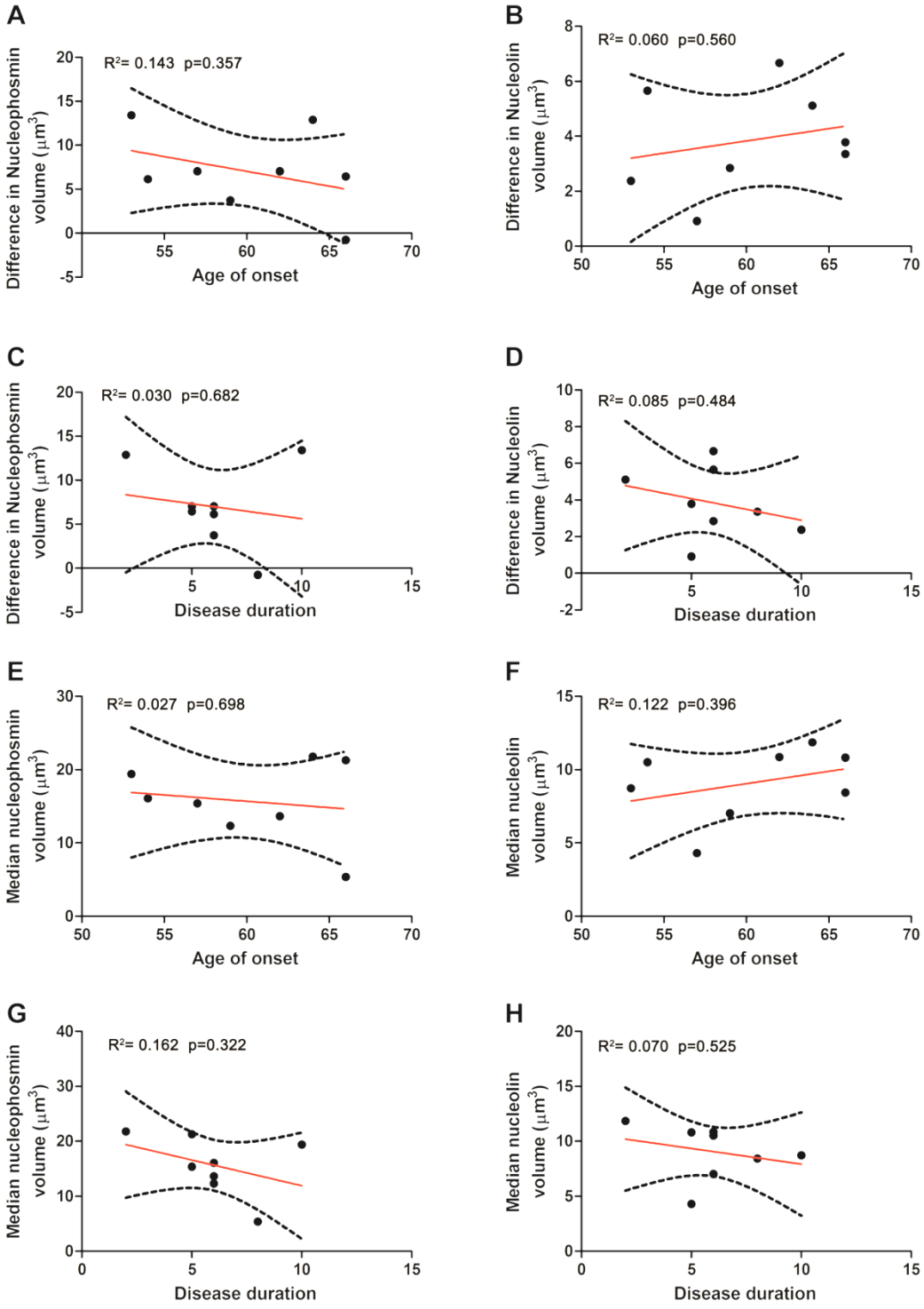


Figure 5.14 Correlation of nucleolar volume with patient disease course

Difference in median nucleolar volume between poly(GR)-inclusion-positive and negative neurons was plotted against patient age of disease onset for (A) nucleophosmin and (B) nucleolin stained sections. Difference in median nucleolar volume between poly(GR)-inclusion-positive and negative neurons was also plotted against patient disease duration for (C) nucleophosmin and (D) nucleolin stained sections. Median nucleolar volume in poly(GR)-inclusion-positive neurons was plotted against patient age of disease onset for (E) nucleophosmin and (F) nucleolin stained sections. Median nucleolar volume in poly(GR)-inclusion-positive was also plotted against patient disease duration for (G) nucleophosmin and (H) nucleolin stained sections. Linear regressions were pereformed on all comparisons. r^2 values are shown as a measure of goodness of fit. p values shown denote the statistical probability of a linear trend with a non-zero slope. Red lines indicate line of best fit. Dashed black lines represent 95% confidence band for line of best fit.

5.3 Discussion

5.3.1 Summary of results

In this chapter, we found that the nucleolar dysfunction reported in *C9orf72* cell models is also evident in *C9orf72* FTD post-mortem patient brains. Furthermore, increased neuronal nucleolar volume was found to be associated with the presence of a poly(GR) inclusion, in support of a toxic role for arginine-containing DPR proteins in *C9orf72* FTD.

We examined nucleolar morphology and poly(GR) aggregate formation in immunostained *C9orf72* FTD and control patient brains by developing a 3D volumetric image analysis protocol. In *C9orf72* FTD patients, neurons containing a poly(GR) inclusion were found to have a larger volume of both nucleophosmin and nucleolin staining compared to poly(GR)-inclusion-negative neurons. *C9orf72* FTD neurons without a poly(GR) inclusion also had a significantly lower volume of nucleophosmin than control patient neurons, however nucleolin volume was similar between these populations. Nucleophosmin was additionally found to occupy a larger proportion of the nucleus in poly(GR)-inclusion-positive neurons than in *C9orf72* FTD neurons without a poly(GR) inclusion. No difference in nuclear volume was observed between control neurons or *C9orf72* FTD neurons with or without a poly(GR) inclusion, suggesting that the observed differences in nucleolar volume are specific to the nucleolus and are not the result of generic cellular swelling.

However, while nucleolar segregation and dispersal of nucleolar proteins has been reported in cell models of *C9orf72*, evidence for these phenomena could not be observed in *C9orf72* post-mortem brain. No significant difference in the volume of diffuse nucleolin outside of a defined nucleolar compartment was observed between control neurons and *C9orf72* FTD neurons with or without a poly(GR) inclusion. Additionally, no difference in the number of distinct nucleolar objects were observed for nucleophosmin or nucleolin staining for any of the groups examined.

Finally, we examined whether the severity of nucleolar phenotype correlates with *C9orf72* FTD patient age of disease onset or disease duration, but found no significant correlation for any of the comparisons tested.

5.3.2 Confirmation of a nucleolar morphological change in post-mortem *C9orf72* FTD patient brain

Contrary to the previous study by Schludi et al. (2015), we observed a significant increase in nucleolar volume in poly(GR)-inclusion-positive neurons compared to poly(GR)-inclusion-negative neurons in *C9orf72* FTD patient frontal cortex. This disparity may be for a number of reasons. Firstly, while the previous study examined two control and two *C9orf72* patient cases, we assessed a larger patient cohort of eight controls, eight heterozygous *C9orf72* FTD patients, and one homozygous *C9orf72* FTD patient, with a larger number of neurons also examined per case, lending greater power to our analysis. Secondly, while the first study examined single plane images and estimated nucleolar volume using feret diameter, we undertook a 3D volumetric approach by taking z-stack images in 20 µm thick brain sections to more accurately assess nucleolar volume. Use of a 2D approach risks erroneous estimation of nucleolar volume if the image plane does not intersect with the widest portion of the nucleolus.

To fully characterise any changes in nucleolar morphology, we measured the volumes of two different nucleolar markers: nucleophosmin, a nuclear chaperone that mainly localises to the granular component of the nucleolus (Frehlick et al., 2007), and nucleolin, a multifunctional protein involved in ribosomal synthesis and maturation that primarily localises to the dense fibrilar centres of the nucleolus (Mongelard and Bouvet, 2007). An increased volume of both of these markers was observed in poly(GR) positive cells. This supports the findings of previous *C9orf72* cell model studies, which have observed increased volume of both nucleolin and nucleophosmin in response to over-expression of

poly(GR) or poly(PR) (Tao et al., 2015; Wen et al., 2014). The increase in nucleolar volume in poly(GR)-inclusion-positive cells in this study is also approximately equal to that seen in studies of *C9orf72* cell models and patient derived cell lines, with nucleolar volume in poly(GR) positive cells found to be between 1.5 and 2 times greater than in poly(GR)-negative cells (Haeusler et al., 2014; Tao et al., 2015; Wen et al., 2014), suggesting a similar effect on nucleoli in *C9orf72* patients compared to cell models. Additionally, we observed a significantly lower volume of nucleophosmin staining in *C9orf72* neurons lacking a poly(GR) inclusion compared to control patient neurons, but the same was not observed for nucleolin volume in these cells. Differential regulation of these two markers could reflect disparate effects of poly(GR) on the granular component and dense fibrilar centres of the nucleolus in *C9orf72* FTD, however further investigation with additional nucleolar markers would be required to determine if this is the case.

5.3.3 Absence of the translocation or fragmentation of nucleolar markers in *C9orf72* FTD post mortem brain

Nucleolar stress is known to promote stabilisation of p53, promoting cell death. This pathway usually involves the disruption of nucleolar integrity and translocation of nucleolar proteins into the nucleoplasm (Boulon et al., 2010). Nucleolin and nucleophosmin were reported to exhibit a more dispersed localisation in *C9orf72* ALS patient B lymphocyte cells compared to control lymphocytes, while nucleolin was also more dispersed in *C9orf72* ALS patient derived iPS neurons and in $(GGGGCC)_{21}$ transfected HEK293T cells compared to controls (Haeusler et al., 2014). In addition, nucleophosmin was observed to translocate into the nucleoplasm in HEK293T cells transfected with $(GR)_{30}$ or $(PR)_{30}$ DNA constructs (Tao et al., 2015). Conversely, in this chapter we did not detect evidence for a change in the volume of diffuse nucleolar proteins outside of a defined nucleolar localisation. No nucleophosmin staining above background signal was detected outside of the nucleolus in

either control or *C9orf72* FTD patient neurons. Nucleolin signal above background fluorescence levels was detected in some cells from both control and *C9orf72* FTD patient neurons, but no significant difference in the volume of this staining was detected between control neurons, *C9orf72* FTD poly(GR)-inclusion-positive neurons or *C9orf72* FTD poly(GR)-inclusion-negative neurons, regardless of whether absolute volume or volume as a proportion of DAPI nuclear volume was measured. However, measuring the volume of diffuse immunostaining assumes a uniform concentration of nucleolin within the measured region, and so does not necessarily reflect the exact amount of nucleolin present outside of the nucleolus in each neuron. A more accurate estimate of the quantity of nucleolin outside of the nucleolus might be obtained by also considering the intensity of the extra-nucleolar signal, which would require a more sophisticated analysis protocol to be designed to quantify this and to account for the patient to patient staining variability. It was also reported that nucleolin exhibited a more "fractured" morphology in B-lymphocytes (Haeusler et al., 2014), suggestive of the nucleolar fragmentation observed in response to some forms of nucleolar stress (Boulon et al., 2010). However, when we examined the number of defined nucleophosmin or nucleolin stained objects per neuron, no significant difference was observed between control or *C9orf72* FTD patient neurons, regardless of the presence or absence of a poly(GR) inclusion, with the majority of neurons containing one nucleolus. Therefore, while we did observe a difference in nucleolar volume, we did not observe nucleolar dispersal or fragmentation in *C9orf72* FTD patient post-mortem cortex in contrast to the observations made in *C9orf72* cell models or patient derived cell lines. This suggests that while current cell models recapitulate the nucleolar dysfunction seen in *C9orf72* FTD patients in some respects, the mechanisms at work in these two systems may still differ, therefore caution should be taken to confirm the relevance of any dysfunction identified in cell models to human disease.

5.3.4 Potential impact of increased nucleolar volume in *C9orf72* FTD

Disruption of normal ribosomal RNA transcription and onset of nucleolar stress has previously been shown to trigger neuronal apoptosis (Kalita et al., 2008; Parlato et al., 2008). Increased nucleolar volume, as was detected in poly(GR) inclusion positive neurons in this chapter, is usually associated with increased ribosomal production and aberrant cell growth, and is observed in cancer cells (Montanaro et al., 2008) as well as in cardiac hypertrophy (Hariharan and Sussman, 2014). Conversely, in neurodegenerative diseases such as Parkinson's Disease (PD) and Alzheimer's disease (AD), nucleolar stress and decline in rRNA synthesis is usually associated with a decrease in nucleolar volume (Mann and Yates, 1982; Mann et al., 1981; Parlato and Kreiner, 2013). Nucleolar shrinkage and decreased nucleolar activity is similarly associated with normal aging (García Moreno et al., 1997). Interestingly, one study found that while nucleolar size is decreased in AD patients in comparison to control patients, it is increased in patients with asymptomatic AD, who exhibit Aβ plaques and neurofibrillary tangles but normal cognitive function (Iacono et al., 2009), suggesting that increased nucleolar volume could represent a protective mechanism. Moreover, while neurons are post-mitotic, larger nucleoli and increased nucleolar activity have been associated with growth with regards to cellular regeneration and outgrowth of axons following injury (Hetman and Pietrzak, 2012). In response to the lack of a correlation observed between DPR inclusion burden and affected brain regions in *C9orf72* FTD/ALS, it has been hypothesised that DPR inclusions could exert a protective role by sequestering soluble DPR species and preventing them from exerting their toxic effect. An increase in nucleolar volume could therefore be interpreted as a mechanism promoting the growth and recovery of cells in which soluble poly(GR) has been sequestered into an inclusion. This theory would also be consistent with the decreased nucleophosmin volume observed in *C9orf72* FTD neurons without a poly(GR) inclusion in comparison to control neurons, if these cells were to harbour a heightened level of soluble poly(GR). However,

while human astrocytes treated with synthetic $(PR)_{20}$ produced increased levels of rRNA and snoRNAs, the levels of mature 18S, 5.8S and 28S ribosomal subunits were decreased (Kwon et al., 2014). A second study also showed that overexpression of $(GR)_{30}$ and $(PR)_{30}$ in HEK293T cells caused increased nucleolar volume but a reduction in mature 18S and 28S transcripts (Tao et al., 2015). Cell models treated with arginine-containing DPR proteins therefore appear to contradict the positive correlation usually observed between nucleolar volume and rate of ribosomal subunit production. Furthermore, decreased maturation of the 28S ribosomal subunit was observed in *C9orf72* ALS patient motor cortex (Haeusler et al., 2014). Increased rRNA production in the context of fewer mature ribosomes may suggest an uncoupling in the chain of ribosomal production and an increase in immature rRNA transcription as an attempted compensation mechanism by the affected neuron. Therefore, the increase in nucleolar volume in *C9orf72* FTD is likely to be pathological and not protective.

While a clear nucleolar enlargement was observed in *C9orf72* poly(GR)-inlcusion-positive neurons, neither the median size of nucleoli in these neurons, nor the difference in median nucleolar size between *C9orf72* poly(GR)-inlcusion-positive and poly(GR)-inclusion-negative neurons were observed to correlate with either patient age of disease onset, or disease duration. It is possible that increasing the statistical power of the analysis by examining nucleolar volume in a larger cohort of patients could reveal the existence of a weak correlation. However, considering the multitude of different disease mechanisms known to be involved in *C9orf72* FTD/ALS, it is likely that nucleolar dysfunction plays a small role in a more complicated overarching process, and therefore a correlation may not be seen due to the dependency of age of onset or duration on a multitude of other factors.

5.3.5 Isolating a causal relationship between poly(GR) inclusion burden and nucleolar dysfunction

While work in this chapter established an association between poly(GR) inclusion presence and increased nucleolar volume, more work is required before a causal relationship can be established between poly(GR) aggregates and nucleolar dysfunction, due to the large number of confounding variables inherently present in patient tissue. As a weak but significant positive correlation was observed between nucleolar volume and nuclear volume, we examined nuclear volume in control, *C9orf72* FTD poly(GR)-inclusion-positive and *C9orf72* FTD poly(GR)-inclusion-negative neurons to ensure that a larger nucleolar volume was not being driven by 1) the presence of poly(GR) inclusions in a sub-population of larger neurons, 2) a generic increase in whole cell size in response to poly(GR) inclusion presence, or 3) greater ease of poly(GR) inclusion detection in larger neurons due to inclusions frequently being larger in these cells. However, no difference in DAPI size was observed, indicating a specific nucleolar enlargement in poly(GR) inclusion positive cells. We also examined nucleolar volume as a proportion of nuclear volume in each neuron. Nucleophosmin was found to occupy a significantly higher proportion of the nucleus in poly(GR)-inclusion-positive neurons compared to negative neurons, however while a trend was observed for a higher nucleolin proportion of the nucleus in poly(GR)-inclusion-positive neurons, this was not found to be significant. This could reflect the very weak correlation between nuclear size and nucleolar volume, meaning that dividing nucleolar volume by DAPI volume would introduce a great deal of noise and therefore limit how informative this comparison is with the current sample size.

It is also important to consider that poly(GR) is known to co-aggregate with other DPR proteins, and therefore further work is needed to determine whether or not these proteins also contribute to nucleolar dysfunction. In conjunction with our work in human post-

mortem brain, we have also been continuing to collaborate with Prof. Linda Partridge's group at the UCL Institute of Heathy Aging and the Max Plank Institute for Biology of Aging to use our protein-only *Drosophila* models to dissect the impact of different DPR proteins on nucleolar morphology. In recent unpublished work, $(GR)_{100}$ or $(GA)_{100}$ repeats were individually expressed using non-GGGGCC, protein-only DNA constructs in *Drosophila* adult neurons under an inducible ELAV-GeneSwitch promoter. *Drosophila* in which the $(GR)_{100}$ or $(GA)_{100}$ transgene had been induced by administration of an RU486 activator were compared against flies that harboured a transgene, but that had received no induction. Nucleoli in poly(GR)-inclusion-positive cells in induced $(GR)_{100}$ expressing flies were found to be significantly larger than those in poly(GR)-negative cells, with an over ten-fold increase in average volume observed. No significant difference was observed in nucleolar volume between poly(GR)-inclusion-negative cells in induced or uninduced $(GR)_{100}$ flies. In contrast, poly(GA)-inclusion-positive cells in induced $(GA)_{100}$ expressing flies underwent a modest but significant 50% increase in volume compared to cells lacking a poly(GA) inclusion. The extensive nucleolar enlargement observed in $(GR)_{100}$ expressing flies, which are known to experience a marked reduction in viability as was detailed in section 4.2.3, strongly implicates poly(GR) in nucleolar dysfunction.

However, more work is still required in *C9orf72* post-mortem brain to compliment work carried out in our *Drosophila* models and determine the relative contributions of different DPR proteins to changes in nucleolar morphology. Furthermore, the small increase in nucleolar volume observed in fly cells bearing a poly(GA) inclusion requires further investigation. To determine the relative contributions of different DPR proteins in human post-mortem brain, co-immunostaining could be carried out using antibodies against poly(GR), a second DPR protein such as poly(GA), and a nucleolar marker. Neurons containing poly(GR)-only inclusions could then be distinguished from those bearing inclusions comprised of both poly(GR) and poly(GA), or poly(GA)-only inclusions, and the

nucleolar volumes could be compared across these populations. It would also be of interest to co-immunostain for poly(PR) inclusions and poly(GR) inclusions, to assess whether both arginine containing dipeptides are similarly associated with an increase in nucleolar volume.

The potential effect of repeat RNA on nucleolar dysfunction should also be taken into account. Both nucleolin and nucleophosmin are known to bind RNA G-quadruplexes, and a previous study associated aborted GGGGCC repeat expansion RNA transcripts with nucleolar dysfunction (Haeusler et al., 2014). In order to examine the contribution of RNA mechanisms of toxicity to nucleolar dysfunction in human *C9orf72* patient brain, RNA FISH could be carried out to detect sense or antisense RNA foci in conjunction with immunofluorescent staining for nucleolar markers to examine how the presence of these different entities might differentially affect nucleolar volume.

Chapter 6: Discussion

6.1 Final Summary

The primary aim of this project was to dissect the different gain-of-function mechanisms in *C9orf72* FTD/ALS, to help understand which pathways are important in disease pathogenesis. To do this, genetic constructs were synthesised to recapitulate the RNA and/or DPR protein gain-of-function aspects of *C9orf72* repeat pathology. The effect of these constructs was then examined in cell culture and *Drosophila* models to understand how these individual components of repeat pathology might initiate disease. This work was coupled with studies in human *C9orf72* FTD post-mortem brain tissue in order to understand whether the effects seen in our model systems are relevant to patients with the disease.

Three types of DNA constructs were generated to produce a genetic toolkit for dissecting gain-of-function mechanisms of pathology: 1) pure repeats which yield both repeat RNA and DPR proteins were generated in a range of lengths from 3 to 103 repeats, 2) RNA-only repeats which produce repeat RNA, but not DPR proteins were generated from 36 to 1152 repeats in length and 3) protein-only repeats that code individually for the five different DPR proteins were produced in lengths of 36 repeats or 100 repeats, with short constructs of 6 or 7 repeats generated with a GFP tag. Together these tools allow the dissection of RNA and protein gain-of-function mechanisms, and the range of repeat lengths generated additionally enables the study of these phenomena in a dose-dependent manner.

These repeat constructs were then transiently transfected into immortalised cell lines to study RNA foci formation and the subcellular localisation of DPR protein aggregates. RNA-only and pure repeat constructs made both sense and antisense foci in cells, and repeat length was found to be positively correlated with the proportion of RNA foci-positive cells,

but not with the number of foci per cell. When protein-only constructs were transfected into cells, poly(GA) of 36 and 100 repeats in length formed star-like cytoplasmic inclusions reminiscent of those seen in *C9orf72* FTD/ALS patient brain with more extensive inclusion formation observed in GA_{100} transfected cells. Short GFP-tagged poly(GR) and poly(PR) peptides and untagged PR_{36} peptides produced in protein-only construct transfected cells were observed to aggregate within the nucleus, while longer PR_{100} protein aggregated in the cytoplasm, suggesting a repeat length dependent effect on poly(PR) localisation. Lastly, poly(AP) and short GFP-tagged poly(GP) peptides were found to exhibit a diffuse localisation throughout the cell.

In collaboration with Prof. Linda Partridge's group, transgenic fly lines were generated using the pure, RNA-only and protein-only repeat constructs that I produced. Pure repeat constructs caused degeneration when expressed in the fly eye and a loss of viability when expressed in adult neurons; however no detectable toxicity was caused by RNA-only constructs, implicating DPR proteins in toxicity (Mizielinska et al., 2014). When protein-only constructs were expressed in flies, arginine-containing DPR proteins poly(GR) and poly(PR) were sufficient to cause severe toxicity, implicating an important toxic role for these molecular species in *C9orf72* FTD/ALS, while poly(GA) also caused a modest reduction in viability.

Human post-mortem *C9orf72* FTD patient brain was then examined to investigate the frequency of inclusion formation by different DPR proteins. To do this, an automated image analysis protocol was designed and tested in the cerebellum. Poly(GA) inclusions were found to be the most frequent, with poly(GP) and (GR) less frequent, and poly(AP) and poly(PR) found only rarely. Some disparity was found in the inclusion burden between the different regions of the cerebellum and between different patients, which calls for further

study to determine what impact DPR protein burden in this area might have on patient disease course.

Finally, as the arginine-containing DPR proteins that were toxic in our *Drosophila* model were reported to bind to nucleoli and cause changes in nucleolar morphology and function in cell model systems (Kwon et al., 2014; May et al., 2014; Schludi et al., 2015; Tao et al., 2015; Wen et al., 2014; Yamakawa et al., 2014; Zu et al., 2013), nucleolar volume was examined in human post mortem *C9orf72* FTD patient brain to determine whether evidence for poly(GR) mediated nucleolar dysfunction could be observed patients with the disease. Neurons in *C9orf72* FTD patients that contained a poly(GR) inclusion were found to have a significantly larger median nucleolar volume compared to neurons in *C9orf72* FTD patients without a poly(GR) inclusion, which is a hallmark of nucleolar stress.

6.2 The role of RNA and DPR protein gain-of-function mechanisms in *C9orf72* FTD/ALS

C9orf72 FTD/ALS is a complex and multifaceted pathology that involves a wide range of different molecular processes, and results in a striking variance in symptoms, pathological subtype and disease onset. This thesis provides some initial experiments and tools to begin to untangle the array of different pathways at work. The results of these experiments implicate a potentially important toxic role for arginine-containing DPR proteins in *C9orf72* FTD/ALS, however this does not necessarily rule out a contribution for RNA gain-of-function mechanisms, toxicity mediated by other DPR proteins, or loss of C9orf72 protein function.

6.2.1 Further investigation of potential RNA gain-of-function disease mechanisms

While RNA-only constructs of up to 288 repeats were not observed to yield a toxic phenotype in our transgenic *Drosophila* model, further experiments are needed to determine whether or not RNA gain-of-function disease mechanisms are important in

patients with *C9orf72* FTD/ALS. Generation of *Drosophila* expressing longer 576 or 1152 RNA-only repeats would be interesting to determine whether a longer repeat length is an important factor in the initiation of repeat RNA mediated toxicity, however the propensity for these longer constructs to undergo deletion due to instability has thus far precluded this option. Expression of pure or RNA-only repeats in an intronic context more akin to that of the repeats found in patients might also be necessary to observe RNA mediated toxicity: sequence surrounding the repeats may affect their expression and pathogenicity, and dispensing with the use of a polyadenylation sequence directly after the repeats would promote their retention within the cell nucleus and therefore allow more extensive RNA foci formation. 160 GGGGCC repeats have previously been expressed within an intronic context in a transgenic *Drosophila* model, which resulted in abundant nuclear sense RNA foci formation, however this had little effect on the survival of these flies (Tran et al., 2015). This suggests that even when expressed within an intronic context, this repeat length is not sufficient to cause RNA-mediated toxicity in flies. The generation of flies expressing antisense orientation RNA-only repeats would also be of interest, as expression of antisense repeat RNA transcripts was absent in the *Drosophila* generated in our study (Mizielinska et al., 2014). Both pure and RNA-only constructs were found to produce the same number of RNA foci and were both similarly able to form a G-quadruplex secondary structure, suggesting that these constructs are equally suitable for modelling RNA gain-of-function mechanisms. However, it might additionally be informative to compare sequestration of RNA-binding proteins in RNA foci formed by pure or RNA-only repeat transcripts in cell culture or *Drosophila* models, in case the presence of the stop codon containing interruptions in the RNA-only repeats disrupts the binding of any proteins crucial to pathogenesis.

When interpreting the lack of RNA toxicity in this model, it is also important to consider that the phylogenic distance between humans and *Drosophila* may mean that some

proteins that bind repeat RNA in human brain and play an important role in disease progression in humans may not be present in flies. Another important next step may therefore be to generate an RNA-only mouse to study RNA gain-of-function mechanisms in a mammalian system (see section 6.4 for further discussion of *C9orf72* mouse models).

Additional analysis of *C9orf72* FTD/ALS patient brain may also assist in our understanding of the role of repeat RNA in pathogenesis. While studies have observed a correlation between RNA foci burden and age of onset in *C9orf72* FTD or ALS patients (Cooper-Knock et al., 2015b; Mizielinska et al., 2013), a more in depth characterisation of the distribution of sense and antisense RNA foci formation in larger patient cohorts, across multiple pathological subtypes, and in a greater range of different brain regions will help to determine how well RNA foci burden correlates with degeneration and disease type. Coupling this type of study with immunofluorescent counterstaining for different RNA-binding proteins to examine variation in sequestration into sense and antisense RNA foci across different brain regions and between patients might also reveal some informative differences. Presently these types of studies are limited by the technical challenges posed by the small size of RNA foci, which are difficult to clearly visualise in human brain and require painstaking analysis by eye in large numbers of high magnification confocal fluorescence z-stack images. Automation of this process would make analysis of much larger data sets possible: this might be done by employing the use of a high throughput confocal fluorescence microscope to collect images, and by designing a more sophisticated image analysis protocol similar to the Definiens protocol used for DPR inclusion quantification in this study. However, even when automated, the image capture process required to generate a large data set of different patients and brain regions counterstained for different RNA-binding proteins would be very time consuming, and analysis of the large number resulting high resolution images would require an exceptional amount of processing power. It may therefore be that a large scale study such as this is not presently

feasible for most individual lab groups with the currently available equipment, but would make an interesting collaborative project.

6.2.2 The consequences of DPR protein expression in model systems and *C9orf72* FTD/ALS patient brain.

The work in this thesis supports a toxic role for DPR proteins within *C9orf72* FTD/ALS patients. The degree of toxicity exhibited by the arginine-containing poly(GR) and poly(PR) DPR proteins in our *Drosophila* model is so severe that for them to be completely benign in human *C9orf72* FTD/ALS patients seems improbable. However, it is true that regional DPR protein inclusion burden does not correlate well with neurodegeneration in *C9orf72* FTD/ALS patients (Davidson et al., 2014; Mackenzie et al., 2015, 2013) and that these aggregates are conspicuously sparse in some important areas of degeneration such as in the spinal cord and motor neurons of *C9orf72* ALS patients (Gomez-Deza et al., 2015). However, it is possible that soluble DPR protein species may instead be responsible for toxicity. Initial studies have identified the presence of soluble poly(GP) and poly(GA) in *C9orf72* FTD/ALS patient brain, which have so far been found to correlate with poly(GP) or poly(GA) inclusion burden in different brain regions (Gendron et al., 2015; Mackenzie et al., 2015). However, analysis of the quantities of soluble poly(GR) and poly(PR) has yet to be carried out. Development of an immunoassay for the detection of soluble poly(GR) and poly(PR) will therefore be important to determine whether soluble levels of these proteins might correlate with the degree of degeneration in different brain regions, despite absence of a correlation between poly(GR) or poly(PR) inclusion burden and degeneration.

It is also true that while poly(GR) and poly(PR) localise to the nucleolus in the cell culture model studied in this thesis, as well as in cell culture models generated by other groups (Kwon et al., 2014; May et al., 2014; Schludi et al., 2015; Tao et al., 2015; Wen et al., 2014; Yamakawa et al., 2014; Zu et al., 2013), within patients these proteins are primarily

detected within star-like inclusions similarly to poly(GA) and p62, or as para-nucleolar inclusions within the nucleus (Schludi et al., 2015), indicating a difference in localisation between cell models and patients. Sequestration of these peptides into inclusions could therefore be predicted to prevent them from interacting with nucleolar factors. However, as the work in this thesis shows an association between poly(GR) inclusion burden and an increase in nucleolar volume in *C9orf72* FTD patient cortex, it may be that neurons with a poly(GR) inclusion also have a high soluble poly(GR) burden, which could be the species that interacts with the nucleolus to cause dysfunction. It may therefore be useful to determine not only the global quantities of soluble poly(GR), but also how soluble levels of this peptide differ in inclusion-bearing and non-inclusion-bearing cells, which would require the microdissection of individual cells and the analysis of the different soluble and insoluble fractions.

There are a number of potential reasons for a difference in DPR protein toxicity in cell culture and *Drosophila* models compared to human *C9orf72* FTD/ALS patients, such as the differences in the genetic context of expression between these systems. The repeat constructs expressed in our transgenic *Drosophila* lines were expressed as a part of poly-adenylated mRNA transcripts, which facilitates the nuclear export and translation of these transcripts (Tran et al., 2015; Weill et al., 2012). Additionally, translation of our protein-only repeats was driven by an ATG start codon allowing canonical translation to occur. These factors may result in a high DPR protein burden in pure repeat and protein-only repeat expressing *Drosophila* in comparison to those produced from the intronic repeats within human *C9orf72* FTD/ALS patients. In patients the accumulation of DPR proteins by the inefficient process of RAN translation is likely to be more gradual, with protein degradation mechanisms potentially serving to reduce levels of these proteins until proteolysis begins to decline in old age (Martinez-Vicente et al., 2005). While slow onset of DPR protein burden in patients will potentially limit their toxicity, these peptides will have a lifetime in

which to gradually inflict damage. Study of patients in the early phases of *C9orf72* FTD/ALS has already shown that DPR protein aggregation pre-dates TDP-43 deposition (Baborie et al., 2014; Proudfoot et al., 2014), implying that these proteins may be the initiators of disease, while TDP-43 mislocalisation may be the final actuator of cell death, or may simply be a consequence of disrupted cell function. An "amyloid-like cascade hypothesis" has been proposed, in which DPR proteins may be responsible for the initiation of disease processes in an analogous fashion to soluble Aβ and amyloid plaques in Alzheimer's disease, which then leads to a snowballing effect of increasing inflammatory response and cell damage that eventually leads to deposition of tau neurofibrillary tangles, or in the case of *C9orf72* FTD/ALS, TDP-43 aggregation (Edbauer and Haass, 2015). Additional detailed quantification of DPR protein aggregation in different brain regions and in different pathological subtypes of early-phase *C9orf72* FTD/ALS patients would help to determine whether DPR proteins do play this type of role in disease initiation, for which the DPR protein aggregate image analysis protocol developed in this thesis could be used. Equally, analysis of soluble species in these patients would also allow us to understand more about the relative roles of insoluble aggregates and soluble monomers or oligomers that may be formed by DPR proteins.

6.2.3 Contribution of different pathological mechanisms to disease heterogeneity

Rather than considering *C9orf72* FTD/ALS to be a pathology for which one key disease causing factor initiates symptom onset, it is likely that all of the molecular mechanisms that arise from the repeat expansion play a role in pathogenesis. The large number of potentially toxic factors caused by the repeat expansion raises the possibility that different mechanisms are more or less important in different patients or brain regions, and thus lead to the degeneration of different cell populations, which might explain the high heterogeneity of this disease.

There are a huge number of different factors potentially affecting disease course in *C9orf72* FTD/ALS patients (summarised in Figure 6.1), which include: 1) differential reduction in the levels of three different *C9orf72* transcripts and two C9orf72 protein isoforms, 2) expression of sense GGGGCC and antisense CCCCGG RNA transcripts, their subsequent aggregation to form sense and antisense RNA foci, the formation of different secondary RNA conformations and the potential sequestration of an array of RNA binding factors, 3) RAN translation of five different DPR proteins which exist in soluble form and as cytoplasmic or nuclear aggregates, some of which may disrupt nucleolar function and nucleocytoplasmic transport (see section 6.3 below), 4) modulation of all of the above processes according to exact repeat length harboured by a particular patient, as well as by other genetic and environmental risk factors. When all of these factors are taken into account, the clinical heterogeneity of this disease seems unsurprising.

Neurons are thought to be particularly susceptible to cellular stress and damage, especially as the brain ages, as they are post-mitotic cells with a high metabolic demand which can result in the build-up of damage caused by physiological stress and excitotoxicity (Saxena and Caroni, 2011). However, which population of neurons are affected first in *C9orf72* FTD/ALS may be dependent on all of the factors listed above, with certain mechanisms or risk factors perhaps favouring damage or TDP-43 accumulation in one cell type over another. One study examined the survival of cultured rat cortical, motor and hippocampal neurons transfected with 42 GGGGCC repeats, and found that cortical and motor neurons exhibited a significant reduction in survival compared to controls, while no significant difference was observed in hippocampal neurons, suggesting that cortical and motor neurons are distinctly sensitive to toxicity mediated by the repeats (Wen et al., 2014). Further cell culture or mouse model studies could be carried out using the RNA-only and protein-only constructs generated in this thesis to determine whether motor neurons or cortical neurons are more susceptible to one gain-of-function mechanism or the other, to

examine if a predominance of one mechanism or the other might promote either an FTD or an ALS disease onset. Some studies in *C9orf72* FTD/ALS patient brain have also highlighted some differences in DPR protein burden in different patients, with a lower poly(GP) burden detected in the cerebellum of *C9orf72* ALS patients compared to FTD or FTD/ALS patients (Gendron et al., 2015), and a higher poly(GA) inclusion burden in the cerebellum of *C9orf72* FTD patients compared to ALS or FTD/ALS patients (Schludi et al., 2015). However these findings were not replicated by another study (Mackenzie et al., 2015), and therefore further research in patient tissue is required to confirm what differences exist between different *C9orf72* disease subtypes.

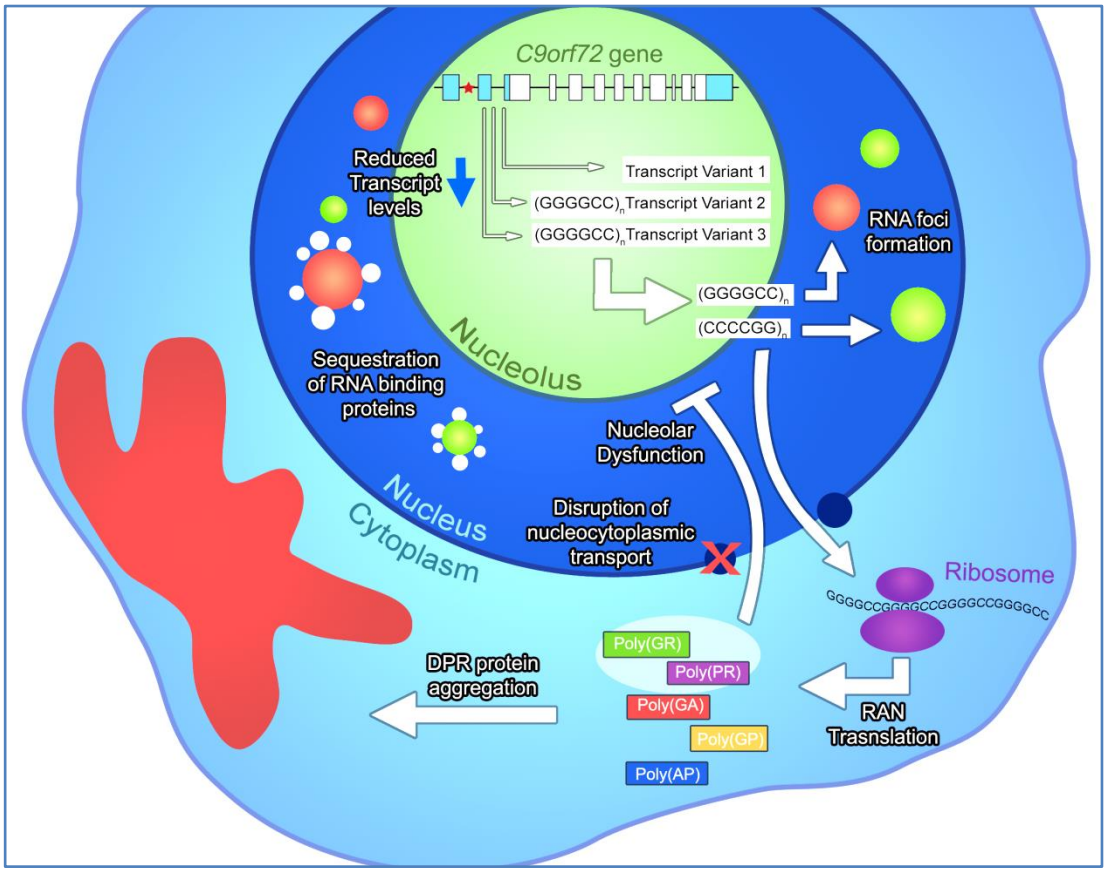


Figure 6.1 *C9orf72* FTD/ALS disease mechanisms.

6.3 Dissecting RNA-mediated and DPR protein-mediated disruption of nucleocytoplasmic transport in *C9orf72* FTD/ALS

Components of the nuclear pore complex (NPC) are known to be extremely long lived in post-mitotic cells, and can become damaged during aging (D'Angelo et al., 2009). Mutations in several nuclear pore components and nucleocytoplasmic transport factors are also associated with nervous system impairment in several rare diseases (Dickmanns et al., 2015). Therefore, the process of nucleocytoplasmic transport may be an important focus for neurodegeneration research. Recently, a number of studies have implicated the disruption of protein and RNA transport between the nucleus and the cytoplasm as a pathological mechanism in *C9orf72* FTD/ALS (Freibaum et al., 2015; Jovičić et al., 2015; Zhang et al., 2015), adding to the number of cellular processes thought to be disrupted by the repeats, however it is unclear what the relative contributions of RNA and DPR protein gain-of-function mechanisms to this process are. The genetic constructs generated in this thesis could therefore also be used to dissect the contributions of RNA and DPR protein gain-of-function mechanisms to the disruption of nucleocytoplasmic transport in *C9orf72* FTD/ALS in future.

6.3.1 Discovery of a GGGGCC repeat mediated disruption of nucleocytoplasmic transport

Three reports were published that undertook genetic modifier screens in transgenic *C9orf72* repeat expressing yeast and fruit fly models to assess whether mutations in other genes causing their gain or loss of function resulted in the suppression or enhancement of repeat toxicity.

One group crossed transgenic flies expressing 30 GGGGCC repeats with other fly lines harbouring mutations in genes previously found to interact with GGGGCC repeat RNA

(Zhang et al., 2015). A second group carried out an unbiased genetic screen in *Drosophila* expressing 58 GGGGCC repeats (Freibaum et al., 2015). Finally, a third group specifically examined the effect of DPR protein expression by carrying out a genetic modifier screen in PR_{50} expressing *Saccharomyces cerevisiae* strains, which were observed to have a dramatically reduced viability.

All three groups found that mutations in genes encoding components of the Ran GTPase nuclear transport cycle modulated toxicity in these different models. Maintenance of a high ratio of GDP-bound Ran to GTP-bound Ran between the cytoplasm and nucleus is critical for normal functioning of nucleocytoplasmic transport. In the cytoplasm, Ran GTPase activating protein (RanGAP) induces Ran to hydrolyse its bound GTP to GDP, while in the nucleus the Ran guanine exchange factor (GEF) RCC1 catalyses the conversion of Ran-GDP to Ran-GTP. A dominant-negative mutation of Ran was found to enhance toxicity in one *Drosophila* model (Freibaum et al., 2015). Additionally, overexpression of RanGAP, promoting nuclear import, suppressed toxicity in the other fly model (Zhang et al., 2015). However, overexpression of RCC1, promoting nuclear export, enhanced toxicity in PR_{50} expressing yeast (Jovičić et al., 2015), suggesting that toxicity in these models is mediated by a deficit in nuclear import. A number of other components of nucleocytoplasmic transport were also found to modulate toxicity, including components of the nuclear pore complex and several karyopherin nuclear transport proteins (Freibaum et al., 2015; Jovičić et al., 2015). Furthermore, the nuclear envelope was observed to have an aberrant wrinkled morphology in flies, and some nucleocytoplasmic transport proteins were additionally observed to form perinuclear aggregates in neurons differentiated from *C9orf72* ALS patient derived iPSCs.

In accordance with these observed defects in nucleocytoplasmic transport machinery, a functional deficit in nuclear import was also observed in these models as well as in *C9orf72*

ALS iPSC derived neurons, with fluorescent marker proteins conjugated to a nuclear localisation signal and a nuclear export signal found to be enriched in the cytoplasm and depleted from the nucleus (Freibaum et al., 2015; Zhang et al., 2015). RNA export was additionally found to be disrupted, with an accumulation of newly synthesised RNA observed in the nuclei of *C9orf72* ALS iPSC derived neurons (Freibaum et al., 2015). These studies therefore implicate the disruption of nuclear protein import and RNA export as potentially important disease mechanisms induced by the *C9orf72* repeat expansion

6.3.2 Determining the cause of nucleocytoplasmic transport defects in *C9orf72* FTD/ALS

The three studies discussed above approached their respective genetic screens with different preconceptions regarding the toxic species at work in *C9orf72* FTD/ALS, but arrived at the same conclusion of the disruption of nucleocytoplasmic transport within this disease. Zhang et al. proposed an RNA-gain-of function mechanism of toxicity, based on the observed interaction of RanGAP with G-quadruplex forming GGGGCC RNA. As RanGAP localises to the cytoplasm and not the nucleus, this would preclude an interaction between RanGAP and nuclear RNA foci, and instead implicate the soluble repeat RNA species or infrequent RNA foci present within the cytoplasmic compartment. While this group were unable to detect DPR proteins within their fly model, this does not preclude their presence at concentrations below the sensitivity threshold of the assay used. Additionally, while the disruption of GGGGCC RNA and G-quadruplex formation was observed to rescue nuclear transport defects, this could equally be predicted to abolish subsequent RAN translation from these repeats. To confirm the involvement of RNA gain-of-function mechanisms, it may be interesting to examine nucleocytoplasmic transport in cell culture or fly models expressing our RNA-only constructs to determine whether a transport defect can still be observed.

Current evidence is more supportive of a DPR protein gain-of-function mechanism mediating a defect in nucleocytoplasmic transport, as Jovicic et al. used protein-only PR_{50} repeats that do not produce repetitive GGGGCC RNA in their yeast and primary neuron studies, which were sufficient to yield a defect in nucleocytoplasmic transport. It would be interesting to also investigate nucleocytoplasmic transport in cell or fly models expressing different protein-only constructs to determine whether DPR proteins other than poly(PR) disrupt nuclear transport. Ascertaining whether or not poly(GR) plays a role in nucleocytoplasmic transport disruption would be of particular importance to determine whether both arginine-containing DPR proteins exert the same effect. Jovicic et al. showed that deletion of many of the same targets that suppressed PR_{50} toxicity also suppress GR_{100} toxicity in yeast, however none of the targets that were effective at suppressing GR_{100} toxicity were confirmed to be involved with nucleocytoplasmic transport, and GR_{100} toxicity was not severe enough in this model to enable the overexpression screen to be carried out. Examination of nucleocytoplasmic transport in cell lines and transgenic fly models expressing poly(GR) in isolation from other DPR proteins or GGGGCC repeat RNA could therefore be carried out to further investigate the involvement of this protein.

Intriguingly, recent research has also found that the short isoform of the C9orf72 protein localises to the nuclear envelope (Xiao et al., 2015), suggesting that a loss-of-function mechanism may also be involved in the disruption of nucleocytoplasmic transport. However, further studies are required to determine whether downregulation of C9orf72 results in a functional nuclear trafficking deficit. Nevertheless, it is fascinating that all three of the proposed mechanisms of repeat expansion pathogenesis appear to converge on the same functional defect.

6.3.3 Nucleocytoplasmic transport defects and protein aggregation

A breakdown in nuclear import could be fundamentally important to *C9orf72* FTD/ALS pathogenesis. Previous studies have shown that knocking down components of the nuclear import machinery in a cell culture model results in the cytoplasmic aggregation of TDP-43 (Nishimura et al., 2010), and disruption of nucleocytoplasmic transport has previously been observed to be associated with TDP-43 aggregation in non-*C9orf72* forms of FTD (Nishimura et al., 2010; Ward et al., 2014), suggesting that this defect could be a point of convergence in different TDP-43 proteinopathies. Indeed, Zhang et al. note that nuclear import defects correlate with aberrant localisation of TDP-43 or its *Drosophila* homolog TBPH in their *C9orf72* ALS iPSC neurons and GGGGCC repeat expressing flies respectively, raising the possibility that TDP-43 aggregation could result from a defect in nucleocytoplasmic transport induced by the *C9orf72* repeat expansion. However a robust causal effect still remains to be demonstrated.

Recent research found that cytoplasmic inclusion formation by a range of different aggregate-prone proteins, including several synthetic proteins with no biological function, inhibit nucleocytoplasmic transport (Woerner et al., 2015), suggesting that it is a general pathological mechanism initiated by aggregating proteins. It may be that DPR protein aggregates also inhibit nucleocytoplasmic transport in the same way. However, cytoplasmic aggregate formation by TDP-43 C-terminal fragments were also specifically found to inhibit nucleocytoplasmic transport in this study (Woerner et al., 2015), which makes it unclear whether the observed defects in nucleocytoplasmic transport in *C9orf72* repeat expansion models are the cause or the consequence of TDP-43 mislocalisation and aggregation. One possibility is that RNA or DPR protein gain-of-function mechanisms of toxicity could cause an initial reduction in nucleocytoplasmic transport, leading to some TDP-43 deposition, which could then produce a positive feedback loop further inhibiting nucleocytoplasmic transport and leading to further TDP-43 and DPR protein aggregation. Interestingly, while

cytoplasmic aggregates caused nucleocytoplasmic defects, nuclear aggregates were not found to disrupt nuclear transport (Woerner et al., 2015), indicating that the compartment that aggregation occurs in is important in determining their pathogenicity. It was noted that in the nuclear compartment these proteins co-immunoprecipitated with the nucleolar protein nucleophosmin, which the authors suggest may chaperone the proteins within the nucleus and block the toxicity of the aggregates (Woerner et al., 2015). This draws some interesting parallels with the findings in this thesis and in other publications which implicate a role for arginine-containing DPR proteins in nucleolar dysfunction. An interesting possibility would be if disruption of normal nucleolar function by poly(GR) and poly(PR) prevents nucleophosmin from carrying out a protective role against nuclear aggregates. However, as Woerner et al. only showed co-localisation of nucleophosmin with aggregating proteins, more functional work is needed to confirm whether it is playing a protective role. One potential experiment could be to examine the effect of nuclear aggregates in a cell culture model with or without the knock-down of nucleophosmin using siRNA to determine whether a difference in aggregate toxicity is observed. However as nucleophosmin is a multifunctional protein, determining whether any increased toxicity caused by nuclear inclusions is due specifically to the loss of nucleophosmin aggregate chaperoning activity or due to a more generic reduction in cell viability caused by a loss of other nucleophosmin functions could be problematic. More work is also required to determine if and how nucleocytoplasmic transport defects and nucleolar dysfunction in repeat expressing cells may be interrelated: this might be investigated by rescuing nucleocytoplasmic transport in a *C*9orf72 cell model or iPSC-derived neurons and examining whether nucleolar morphology is also rescued as a result, or conversely rescuing nucleolar function and examining whether nucleocytoplasmic transport is restored.

6.4 Mouse models for *C9orf72* FTD/ALS

While modelling disease in cell culture systems and simple organisms such as fruit fly, nematode worm and zebrafish provide useful tools for understanding basic pathological mechanisms, transgenic mouse models allow disease to be studied in a more complex mammalian nervous system and facilitate more sophisticated behavioural tests. This not only makes a more detailed dissection of disease mechanisms possible, but also allows preclinical testing of promising therapeutics. Development of *C9orf72* FTD/ALS mouse models that mirror disease progression observed in human patients is therefore important.

6.4.1 Current *C9orf72* repeat expansion mouse models

A number of *C9orf72* loss-of-function and gain-of-function mouse models have been generated to date. In the loss-of-function models produced so far, no neurodegenerative features have been observed. Treatment of mice with antisense oligonucleotides against *C9orf72*, causing a reduction in transcript levels, was not found to cause any signs of neurodegeneration or behavioural changes (Lagier-Tourenne et al., 2013). More recently, *Nestin-Cre*$^{+/-}$*;C9orf72*$^{fl/fl}$ mice were generated with a conditional knock-out of *C9orf72* in neurons and glia. These mice exhibit a reduced body weight, but no reduction in viability, neurodegeneration or motor defects (Koppers et al., 2015). A deficit of C9orf72 protein therefore does not appear to cause any phenotypes characteristic of *C9orf72* FTD/ALS.

The first published gain-of-function mouse model for *C9orf72* FTD/ALS expresses 80 GGGGCC repeats under an inducible tetracycline promotor (Hukema et al., 2014). These mice exhibit ubiquitin positive inclusions, but neither poly(GA) inclusions or TDP-43 inclusions could be detected after 12 weeks of doxycycline treatment. Additionally, the authors do not report the presence of any RNA foci, and no behavioural deficits or neurodegeneration were observed, therefore the recapitulation of *C9orf72* FTD/ALS pathology by this mouse is limited.

A mouse with 66 GGGGCC repeats overexpressed in the CNS by adeno-associated virus (AAV) mediated transduction was the first rodent model for *C9orf72* FTD/ALS that recapitulated elements of repeat RNA and DPR protein pathology (Chew et al., 2015). Sense RNA foci and sense DPR proteins poly(GA), poly(GP) and poly(GR) were detected in the brain and spinal cord of these mice at 6 months of age, but the presence of antisense RNA foci, or antisense DPR protein aggregates was not investigated. TDP-43 was also found to mislocalise in these mice, demonstrating that repeat gain-of-function in the absence of a C9orf72 protein loss-of-function can induce this important hallmark of ALS and FTD. These mice additionally exhibited neuronal loss in the cortex and cerebellum, along with motor defects and behavioural problems such as anxiety, antisocial behaviour and hyperactivity, drawing parallels with some of the symptoms observed in FTD and ALS patients (Chew et al., 2015). However, two more recent transgenic mouse models expressing much longer repeat expansions using a bacterial artificial chromosome (BAC) were not able to replicate the functional deficits or neurodegeneration observed in this model (O'Rourke et al., 2015; Peters et al., 2015), indicating that the overexpression in the $(GGGGCC)_{66}$ mouse may be driving the phenotypes seen. One of these models expressed the first six exons of the human *C9orf72* gene containing approximately 500 GGGGCC repeats within the first intron and around 140 kb of upstream genomic sequence (Peters et al., 2015). Mice exhibited sense and antisense RNA foci in neurons as well as both soluble poly(GP) and poly(GP) inclusions, however, the presence or absence of other DPR proteins was not reported. No reduction in survival, and no behavioural or cognitive deficits were noted in repeat expressing mice comparison to non-transgenic littermates. Furthermore, no TDP-43 pathology was observed. The second BAC transgenic mice exhibit an array of transgenes with between 100 and 1000 hexanucleotide repeats within the first intron of a complete human *C9orf72* gene (O'Rourke et al., 2015). Again both sense and antisense RNA foci were observed in neurons, as well as both soluble poly(GP) and poly(GP) inclusions, which were

found to be most abundant in the cerebellum and least frequent in the spinal cord, mirroring the distribution pattern observed in *C9orf72* FTD/ALS patients (Schludi et al., 2015). However, the presence or absence of other DPR proteins was again not reported (O'Rourke et al., 2015). This group also observed an increase in neuronal nucleolin staining area in repeat expansion expressing mice, in accord with the similar increase in nucleolar volume observed in *C9orf72* FTD/ALS patients in this thesis, suggesting that these mice successfully model this aspect of the disease. Conversely, this group did not detect a difference in the levels of any ribosomal RNA in 8 month old mice, suggesting this morphological change in nucleoli is not linked to a nucleolar dysfunction; however it may be of interest to examine older mice to determine whether a reduction in ribosomal RNA maturation develops with age. Once more, no neurodegeneration, behavioural defects or reduction in survival were noted in this model, with no TDP-43 pathology observed. Interestingly, no sequestration of Pur α, hnRNP A3, hnRNP A2/ B1, or hnRNP H was observed by RNA foci in these animals, as had previously been noted in human *C9orf72* FTD/ALS patient tissue, bringing into question what factors might affect binding of these proteins to the repeat RNA (O'Rourke et al., 2015).

6.4.2 Future research prospects for *C9orf72* FTD/ALS mouse models

The lack of neurodegeneration in the current *C9orf72* BAC transgenic mouse models despite the presence of RNA foci and DPR protein aggregate formation is very interesting, especially in the light of reports from recent conferences that other soon to be published gain-of-function and loss-of-function *C9orf72* mouse models do exhibit neurodegeneration and a reduction in survival (Dance, 2015). Examination of differences in in these mice may help to unravel what factors lead from the presence of the histopathological hallmarks of disease to the onset of neurodegeneration. For example, upstream or downstream sequences surrounding the *C9orf72* gene may affect the pathogenicity of the repeats in some way. The regional or cell type expression pattern of the transgene in various mouse

models may also affect disease onset, as well as the expression level of the repeat expansion. Furthermore, the BAC mouse models published so far express part or the entire human *C9orf72* gene in addition to endogenous levels of the mouse *C9orf72* homolog: it would therefore be interesting to determine whether overexpression of C9orf72 protein is protective in these models, or whether a partial loss of C9orf72 function is required for disease onset. To determine this, these gain-of-function mouse models could perhaps be crossed with the existing mice with a conditional *C9orf72* knock-out in neurons and glia (Koppers et al., 2015). While neither these loss-of-function mice or the gain-of-function BAC transgenic mice individually show a neurodegenerative phenotype, it is possible that a phenotype could be observed in the resulting cross.

There are also clearly many differences between mice and patients that could affect the toxicity of the repeats. Mice do not undergo spontaneous neurodegeneration in old age as humans and some higher mammals do and therefore may not be as susceptible to neurodegeneration. This could be one reason that a number of existing neurodegenerative mouse models fail to accurately recapitulate the symptoms of human disease despite the expression of disease relevant mutations. Lifespan may be an important factor as disease onset in *C9orf72* FTD/ALS patients most often only occurs after many decades, while mice only live for up to two years. In a non-overexpression system this may not give the repeat expansion sufficient time to cause neurodegeneration in a mouse during its shorter life time.

Due to the lack of complete penetrance and the high variability in disease onset seen in human *C9orf72* expansion carriers, a “two hit hypothesis” has also been suggested in which one or more additional environmental stressors or genetic risk factors are required in combination with the presence of the repeat expansion in order to initiate disease (Edbauer and Haass, 2015), which may not be present in inbred mouse lines housed in a

very safe and artificial environment mostly free from stress. It might therefore be interesting to investigate whether introducing pharmacological or environmental stressors to these mouse models induces a neurodegenerative phenotype.

The RNA-only and protein-only DNA constructs developed in this thesis would be interesting candidates for generation of *C9orf72* mouse models. In theory, expression of our stop codon containing RNA-only repeats could be used to make a mouse that produces sense and antisense RNA foci, but not DPR proteins. This could be compared with mice made to express our protein-only repeats coding individually for each of the different DPR proteins to examine whether a difference in phenotype can be observed between mice in an analogous fashion to the observations made in our *Drosophila* studies (Mizielinska et al., 2014). It would be of great interest to determine whether poly(GR) or poly(PR) are sufficient to cause toxicity in a mammalian system, and whether any RNA mediated toxicity could be observed in an RNA-only mouse, which would indicate an important biological difference between these two organisms. Nucleolar morphology and function could also be studied in RNA-only and protein-only mice to clarify whether poly(GR) or poly(PR) expression is necessary and sufficient to cause nucleolar dysfunction.

6.5 Treatment prospects for *C9orf72* FTD/ALS

6.5.1 Development of potential therapies for *C9orf72* FTD/ALS

A better understanding of the pathological mechanisms at work in *C9orf72* FTD/ALS will hopefully yield new treatment prospects, which are sorely needed for these patients. One promising treatment option is the use of antisense oligonucleotides (ASOs), short nucleic acid sequences that bind to complimentary RNA transcripts within the cell, which either sterically inhibit their interaction with other binding partners, or promote their degradation by RNase H (Dias and Stein, 2002). ASOs have previously been shown to reduce RNA foci formation and normalise aberrant splicing in a mouse model of myotonic dystrophy

(Mulders et al., 2009), and are now being tested in models of *C9orf72* FTD/ALS. Several groups administered ASOs targeted against the *C9orf72* locus in neurons differentiated from *C9orf72* ALS patient derived iPSCs, which were found to reduce sense RNA foci formation and partially reversed transcriptome changes observed in *C9orf72* iPSCs (Donnelly et al., 2013; Sareen et al., 2013). One study found that ASOs prevented the induction of glutamate excitotoxicity in *C9orf72* iPSC-derived neurons, and reduced the interaction between sense RNA foci and the RNA binding protein ADARB2 (Donnelly et al., 2013). Additionally, ASO treatment restored the disrupted Ran gradient between the nucleus and cytoplasm in *C9orf72* ALS iPSC-derived neurons, and rescued the nuclear mis-localisation of TDP-43 (Zhang et al., 2015). In another study, ASOs targeted similarly against the *C9orf72* locus were administered to fibroblasts derived from *C9orf72* ALS patients and were found to reduce sense RNA foci formation, however did not correct the aberrant RNA transcript expression profile in these cells (Lagier-Tourenne et al., 2013). This may be due to the influence of antisense RNA transcripts within *C9orf72* FTD/ALS patient cells, which may need to be independently targeted with a second ASO in order to substantively reverse transcript changes in patients.

Bioactive small molecules have also been designed that bind to GGGGCC repeats, and affect secondary structure formation. One study found a particular compound that binds GGGGCC repeats reduced RNA foci formation and poly(GP) inclusion formation in neurons differentiated from *C9orf72* patient fibroblasts, with no significant effect on *C9orf72* transcript levels detected (Su et al., 2014). The porphyrin molecule TMPyP4, which is known to bind and distort RNA G-quadruplexes (Morris et al., 2012), was also found to result in the disruption of GGGGCC RNA binding to the proteins hnRNPA1 and SF2 (Zamiri et al., 2014). These small molecule compounds could therefore also be developed as new therapeutic interventions for *C9orf72* FTD/ALS.

While the above ASOs and small molecules target RNA transcripts and therefore disrupt RNA gain-of-function disease mechanisms, they may also affect downstream RAN translation of the transcripts and therefore simultaneously disrupt DPR protein gain-of-function disease mechanisms. This could be beneficial if both RNA gain-of-function and DPR protein gain-of-function disease mechanisms are eventually found to be important to the progression of *C9orf72* FTD/ALS. One study found that immunostaining of poly(GP) was not changed in *C9orf72* iPSC-derived neurons before or after ASO treatment despite disruption of sense RNA foci formation (Donnelly et al., 2013), however this could be attributed to a slow turnover of poly(GP), so additional characterisation may be required. As other studies investigating ASO treatment have not examined the effect of these molecules on RAN translation (Lagier-Tourenne et al., 2013; Sareen et al., 2013), it remains to be determined how this treatment affects DPR protein expression.

As ASOs and RNA targeted small molecules may result in the reduction or loss of normal C9orf72 protein function due to degradation or steric hindrance of the *C9orf72* RNA transcripts, the potential of a detrimental C9orf72 loss-of-function effect in patients should be investigated. While it was found that use of ASOs to reduce *C9orf72* transcript levels in mice for 18 weeks was well tolerated and did not produce any characteristic behavioural or pathological features characteristic of neurodegeneration (Lagier-Tourenne et al., 2013), it is not yet understood how loss-of-function effects might contribute to the disease in humans, and whether this may exacerbate symptoms. Future treatment development therefore should aim to minimise the impact these molecules have on native C9orf72 protein in *C9orf72* FTD/ALS patients.

While directly targeting *C9orf72* repeats may present a promising strategy for treating patients with this mutation, it would not be effective for patients with other sporadic or familial forms of FTD or ALS. Future research into the disease mechanisms involved in

C9orf72 FTD/ALS and other forms of FTD or ALS may start to uncover points of convergence in these pathologies that can then be targeted in order to develop interventions for a much larger number of patients.

6.5.2 Biomarkers for *C9orf72* FTD/ALS

The development of effective treatments for FTD and ALS patients not only requires the identification of suitable targets for pharmacological interventions, but also early disease detection in patients, and effective monitoring of treatment efficacy in clinical trials. To achieve this, clear biomarkers are required to indicate the disease progression in a patient, which can take the form of changes in brain imaging, or in a molecular profile in patient biological samples. The presence of the *C9orf72* repeat expansion can be detected in patients using techniques such as repeat-primed PCR or Southern blotting against the repeats in patient tissue samples, which will allow patients with a family history of FTD or ALS to be screened and treated early in the disease course, providing an opportunity to halt disease progression before extensive neurodegeneration occurs.

To determine the extent of disease progression in a patient and the efficacy of a treatment intervention, several markers are available. Brain imaging techniques such as magnetic resonance imaging (MRI) or positron emission tomography (PET) can be used to assess the degree and pattern of cortical or subcortical grey matter loss. RNA foci formation can also be assessed in fibroblast or blood samples from patients with the use of fluorescence *in situ* hybridisation, which may provide a read out of the efficacy of system wide interventions. To give a readout of disease progression in the nervous system, a number of secreted molecular disease indicators within the CNS are also being investigated: one group found that poly(GP) protein can be detected by ELISA assay in the cerebrospinal fluid (CSF) (Su et al., 2014). An immunoassay for the detection of poly(GA) has also been developed (Gendron et al., 2015; van Blitterswijk et al., 2015) which could be used to determine

whether changes in the level of this DPR protein can also be detected in the CSF of *C9orf72* FTD/ALS patients. Additional potentially useful biomarkers have been highlighted by another group in the form of secreted CNS proteins found to be upregulated in *C9orf72* ALS iPSC-derived neurons (Donnelly et al., 2013), however further investigation is required to confirm whether the upregulation of these markers can be detected in human patients. As work in this thesis implicates arginine-containing DPR proteins as toxic species in *C9orf72* FTD/ALS pathology, these proteins could provide a helpful marker of disease progression, therefore future development of immunoassays for quantifying the levels of poly(GR) and poly(PR) in patient CSF may be beneficial.

6.6 Conclusion

The discovery of the *C9orf72* repeat expansion has revolutionised our understanding of both ALS and FTD, and has provided a host of interesting new research avenues that will not only enhance our understanding of disease course in patients with the repeat expansion, but also of all ALS and FTD cases, and of neurodegeneration in general. This project focussed on dissecting RNA and DPR protein gain-of-function mechanisms in *C9orf72* FTD/ALS, and has highlighted arginine-containing DPR protein toxicity as a potentially important mechanism in *C9orf72* FTD/ALS, which could be mediated in part by disruption of nucleolar function. A range of different DNA constructs and image analysis tools have also been produced in this work that can be adapted and repurposed in future for further dissection of RNA and DPR protein gain-of-function mechanisms in model organisms and in patient tissue. A great number of questions still remain to be answered about the pathological mechanisms at work in *C9orf72* FTD/ALS and new discoveries often seem to lead to more questions than answers. However, in this fast moving field, breakthroughs in our understanding are sure to be imminent, moving us closer towards the ultimate goal of developing effective treatments for ALS and FTD.

References

Ahmed, Z., Mackenzie, I.R.A., Hutton, M.L., Dickson, D.W., 2007. Progranulin in frontotemporal lobar degeneration and neuroinflammation. J. Neuroinflammation 4, 7. doi:10.1186/1742-2094-4-7

Al-Mahdawi, S., Pinto, R.M., Ismail, O., Varshney, D., Lymperi, S., Sandi, C., Trabzuni, D., Pook, M., 2008. The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. Hum. Mol. Genet. 17, 735–46. doi:10.1093/hmg/ddm346

Almeida, S., Gascon, E., Tran, H., Chou, H.J., Gendron, T.F., Degroot, S., Tapper, A.R., Sellier, C., Charlet-Berguerand, N., Karydas, A., Seeley, W.W., Boxer, A.L., Petrucelli, L., Miller, B.L., Gao, F.-B., 2013. Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol. 126, 385–99. doi:10.1007/s00401-013-1149-y

Alonso, A., Logroscino, G., Hernán, M.A., 2010. Smoking and the risk of amyotrophic lateral sclerosis: a systematic review and meta-analysis. J. Neurol. Neurosurg. Psychiatry 81, 1249–52. doi:10.1136/jnnp.2009.180232

Al-Sarraj, S., King, A., Troakes, C., 2011. p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD. Acta … 122, 691–702. doi:10.1007/s00401-011-0911-2

Ash, P.E.A., Bieniek, K.F., Gendron, T.F., Caulfield, T., Lin, W.-L., Dejesus-Hernandez, M., van Blitterswijk, M.M., Jansen-West, K., Paul, J.W. 3rd, Rademakers, R., Boylan, K.B., Dickson, D.W., Petrucelli, L., 2013. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron 77, 639–646. doi:10.1016/j.neuron.2013.02.004

Ayala, Y.M., Misteli, T., Baralle, F.E., 2008. TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression. Proc. Natl. Acad. Sci. 105, 3785–3789. doi:10.1073/pnas.0800546105

Baborie, A., Griffiths, T.D., Jaros, E., Perry, R., McKeith, I.G., Burn, D.J., Masuda-Suzukake, M., Hasegawa, M., Rollinson, S., Pickering-Brown, S., Robinson, A.C., Davidson, Y.S., Mann, D.M., 2014. Accumulation of dipeptide repeat proteins predates that of TDP-43 in Frontotemporal Lobar Degeneration associated with hexanucleotide repeat expansions in C9ORF72 gene. Neuropathol. Appl. Neurobiol. doi:10.1111/nan.12178

Baker, M., Mackenzie, I.R., Pickering-Brown, S.M., Gass, J., Rademakers, R., Lindholm, C., Snowden, J., Adamson, J., Sadovnick, A.D., Rollinson, S., Cannon, A., Dwosh, E., Neary, D., Melquist, S., Richardson, A., Dickson, D., Berger, Z., Eriksen, J., Robinson, T., Zehr, C., Dickey, C.A., Crook, R., McGowan, E., Mann, D., Boeve, B., Feldman, H., Hutton, M.,

2006. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. Nature 442, 916–9. doi:10.1038/nature05016

Beck, J., Poulter, M., Hensman, D., Rohrer, J.D., Mahoney, C.J., Adamson, G., Campbell, T., Uphill, J., Borg, A., Fratta, P., Orrell, R.W., Malaspina, A., Rowe, J., Brown, J., Hodges, J., Sidle, K., Polke, J.M., Houlden, H., Schott, J.M., Fox, N.C., Rossor, M.N., Tabrizi, S.J., Isaacs, A.M., Hardy, J., Warren, J.D., Collinge, J., Mead, S., 2013. Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. Am. J. Hum. Genet. 92, 345–53. doi:10.1016/j.ajhg.2013.01.011

Bede, P., Bokde, A.L.W., Byrne, S., Elamin, M., McLaughlin, R.L., Kenna, K., Fagan, A.J., Pender, N., Bradley, D.G., Hardiman, O., 2013. Multiparametric MRI study of ALS stratified for the C9orf72 genotype. Neurology 81, 361–9. doi:10.1212/WNL.0b013e31829c5eee

Bell, M.V., Hirst, M.C., Nakahori, Y., MacKinnon, R.N., Roche, A., Flint, T.J., Jacobs, P.A., Tommerup, N., Tranebjaerg, L., Froster-Iskenius, U., Kerr, B., Turner, G., Lindenbaum, R.H., Winter, R., Prembrey, M., Thibodeau, S., Davies, K.E., 1991. Physical mapping across the fragile X: Hypermethylation and clinical expression of the fragile X syndrome. Cell 64, 861–866. doi:10.1016/0092-8674(91)90514-Y

Belzil, V. V, Bauer, P.O., Prudencio, M., Gendron, T.F., Stetler, C.T., Yan, I.K., Pregent, L., Daughrity, L., Baker, M.C., Rademakers, R., Boylan, K., Patel, T.C., Dickson, D.W., Petrucelli, L., 2013. Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. Acta Neuropathol. 126, 895–905. doi:10.1007/s00401-013-1199-1

Biedler, J.L., Helson, L., Spengler, B.A., 1973. Morphology and growth, tumorigenicity, and cytogenetics of human neuroblastoma cells in continuous culture. Cancer Res. 33, 2643–52.

Bochman, M.L., Paeschke, K., Zakian, V. a., 2012. DNA secondary structures: stability and function of G-quadruplex structures. Nat. Rev. Genet. 13, 770–780. doi:10.1038/nrg3296

Boeve, B.F., Boylan, K.B., Graff-Radford, N.R., DeJesus-Hernandez, M., Knopman, D.S., Pedraza, O., Vemuri, P., Jones, D., Lowe, V., Murray, M.E., Dickson, D.W., Josephs, K.A., Rush, B.K., Machulda, M.M., Fields, J.A., Ferman, T.J., Baker, M., Rutherford, N.J., Adamson, J., Wszolek, Z.K., Adeli, A., Savica, R., Boot, B., Kuntz, K.M., Gavrilova, R., Reeves, A., Whitwell, J., Kantarci, K., Jack, C.R., Parisi, J.E., Lucas, J.A., Petersen, R.C., Rademakers, R., 2012. Characterization of frontotemporal dementia and/or amyotrophic lateral sclerosis associated with the GGGGCC repeat expansion in C9ORF72. Brain 135, 765–83. doi:10.1093/brain/aws004

Boisvert, F.-M., van Koningsbruggen, S., Navascués, J., Lamond, A.I., 2007. The multifunctional nucleolus. Nat. Rev. Mol. Cell Biol. 8, 574–585. doi:10.1038/nrm2184

Bomsztyk, K., Denisenko, O., Ostrowski, J., 2004. hnRNP K: one protein multiple processes. Bioessays 26, 629–38. doi:10.1002/bies.20048

Boulon, S., Westman, B.J., Hutten, S., Boisvert, F.-M., Lamond, A.I., 2010. The nucleolus under stress. Mol. Cell 40, 216–27. doi:10.1016/j.molcel.2010.09.024

Boxer, A.L., Mackenzie, I.R., Boeve, B.F., Baker, M., Seeley, W.W., Crook, R., Feldman, H., Hsiung, G.-Y.R., Rutherford, N., Laluz, V., Whitwell, J., Foti, D., McDade, E., Molano, J., Karydas, A., Wojtas, A., Goldman, J., Mirsky, J., Sengdy, P., Dearmond, S., Miller, B.L., Rademakers, R., 2011. Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. J. Neurol. Neurosurg. Psychiatry 82, 196–203. doi:10.1136/jnnp.2009.204081

Bradley, W.G., Mash, D.C., 2009. Beyond Guam: the cyanobacteria/BMAA hypothesis of the cause of ALS and other neurodegenerative diseases. Amyotroph. Lateral Scler. 10 Suppl 2, 7–20. doi:10.3109/17482960903286009

Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.-P., Hudson, T., Sohn, R., Zemelman, B., Snell, R.G., Rundle, S.A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P.S., Shaw, D.J., Housman, D.E., 1992. Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell 68, 799–808. doi:10.1016/0092-8674(92)90154-5

Burrell, J.R., Kiernan, M.C., Vucic, S., Hodges, J.R., 2011. Motor neuron dysfunction in frontotemporal dementia. Brain 134, 2582–94. doi:10.1093/brain/awr195

Byrne, S., Elamin, M., Bede, P., Shatunov, A., Walsh, C., Corr, B., Heverin, M., Jordan, N., Kenna, K., Lynch, C., McLaughlin, R.L., Iyer, P.M., O'Brien, C., Phukan, J., Wynne, B., Bokde, A.L., Bradley, D.G., Pender, N., Al-Chalabi, A., Hardiman, O., 2012. Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a C9orf72 repeat expansion: a population-based cohort study. Lancet Neurol. 11, 232–240. doi:10.1016/S1474-4422(12)70014-5

Byrne, S., Heverin, M., Elamin, M., Walsh, C., Hardiman, O., 2014. Intermediate repeat expansion length in C9orf72 may be pathological in amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Frontotemporal Degener. 15, 148–50. doi:10.3109/21678421.2013.838586

Chen, C.X., Cho, D.S., Wang, Q., Lai, F., Carter, K.C., Nishikura, K., 2000. A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. RNA 6, 755–67.

Chen, Y.-Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A., Nicholson, G.A., Auer-Grumbach, M., Wagner, K., De Jonghe, P., Griffin, J.W., Fischbeck, K.H., Timmerman, V., Cornblath, D.R., Chance, P.F., 2004. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). Am. J. Hum. Genet. 74, 1128–35. doi:10.1086/421054

Chew, J., Gendron, T.F., Prudencio, M., Sasaguri, H., Zhang, Y.-J., Castanedes-Casey, M., Lee, C.W., Jansen-West, K., Kurti, A., Murray, M.E., Bieniek, K.F., Bauer, P.O., Whitelaw, E.C., Rousseau, L., Stankowski, J.N., Stetler, C., Daughrity, L.D., Perkerson, E.A., Desaro, P., Johnston, A., Overstreet, K., Edbauer, D., Rademakers, R., Boylan, K.B., Dickson, D.W., Fryer, J.D., Petrucelli, L., 2015. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. Science (80-.). 348, 1151–1154. doi:10.1126/science.aaa9344

Chow, C.Y., Landers, J.E., Bergren, S.K., Sapp, P.C., Grant, A.E., Jones, J.M., Everett, L., Lenk, G.M., McKenna-Yasek, D.M., Weisman, L.S., Figlewicz, D., Brown, R.H., Meisler, M.H., 2009. Deleterious variants of FIG4, a phosphoinositide phosphatase, in patients with ALS. Am. J. Hum. Genet. 84, 85–8. doi:10.1016/j.ajhg.2008.12.010

Ciura, S., Lattante, S., Le Ber, I., Latouche, M., Tostivint, H., Brice, A., Kabashi, E., 2013. Loss of function of C9orf72 causes motor deficits in a zebrafish model of Amyotrophic Lateral Sclerosis. Ann. Neurol. doi:10.1002/ana.23946

Cooper-Knock, J., Bury, J.J., Heath, P.R., Wyles, M., Higginbottom, A., Gelsthorpe, C., Highley, J.R., Hautbergue, G., Rattray, M., Kirby, J., Shaw, P.J., 2015a. C9ORF72 GGGGCC Expanded Repeats Produce Splicing Dysregulation which Correlates with Disease Severity in Amyotrophic Lateral Sclerosis. PLoS One 10, e0127376. doi:10.1371/journal.pone.0127376

Cooper-Knock, J., Hewitt, C., Highley, J.R., Brockington, A., Milano, A., Man, S., Martindale, J., Hartley, J., Walsh, T., Gelsthorpe, C., Baxter, L., Forster, G., Fox, M., Bury, J., Mok, K., McDermott, C.J., Traynor, B.J., Kirby, J., Wharton, S.B., Ince, P.G., Hardy, J., Shaw, P.J., 2012. Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72. Brain 135, 751–64. doi:10.1093/brain/awr365

Cooper-Knock, J., Higginbottom, A., Stopford, M.J., Highley, J.R., Ince, P.G., Wharton, S.B., Pickering-Brown, S., Kirby, J., Hautbergue, G.M., Shaw, P.J., 2015b. Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. Acta Neuropathol. 130, 63–75. doi:10.1007/s00401-015-1429-9

Cooper-Knock, J., Walsh, M.J., Higginbottom, A., Robin Highley, J., Dickman, M.J., Edbauer, D., Ince, P.G., Wharton, S.B., Wilson, S.A., Kirby, J., Hautbergue, G.M., Shaw, P.J., 2014. Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain 137, 2040–51. doi:10.1093/brain/awu120

Couthouis, J., Hart, M.P., Shorter, J., DeJesus-Hernandez, M., Erion, R., Oristano, R., Liu, A.X., Ramos, D., Jethava, N., Hosangadi, D., Epstein, J., Chiang, A., Diaz, Z., Nakaya, T., Ibrahim, F., Kim, H.-J., Solski, J.A., Williams, K.L., Mojsilovic-Petrovic, J., Ingre, C., Boylan, K., Graff-Radford, N.R., Dickson, D.W., Clay-Falcone, D., Elman, L., McCluskey, L., Greene, R., Kalb, R.G., Lee, V.M.-Y., Trojanowski, J.Q., Ludolph, A., Robberecht, W., Andersen, P.M., Nicholson, G.A., Blair, I.P., King, O.D., Bonini, N.M., Van Deerlin, V., Rademakers, R., Mourelatos, Z., Gitler, A.D., 2011. A yeast functional screen predicts new candidate ALS disease genes. Proc. Natl. Acad. Sci. U. S. A. 108, 20881–90.

doi:10.1073/pnas.1109434108

Cruts, M., Gijselinck, I., van der Zee, J., Engelborghs, S., Wils, H., Pirici, D., Rademakers, R., Vandenberghe, R., Dermaut, B., Martin, J.-J., van Duijn, C., Peeters, K., Sciot, R., Santens, P., De Pooter, T., Mattheijssens, M., Van den Broeck, M., Cuijt, I., Vennekens, K., De Deyn, P.P., Kumar-Singh, S., Van Broeckhoven, C., 2006. Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. Nature 442, 920–4. doi:10.1038/nature05017

Cruts, M., Gijselinck, I., Van Langenhove, T., van der Zee, J., Van Broeckhoven, C., 2013. Current insights into the C9orf72 repeat expansion diseases of the FTLD/ALS spectrum. Trends Neurosci. 36, 450–9. doi:10.1016/j.tins.2013.04.010

D'Angelo, M.A., Raices, M., Panowski, S.H., Hetzer, M.W., 2009. Age-dependent deterioration of nuclear pore complexes causes a loss of nuclear integrity in postmitotic cells. Cell 136, 284–95. doi:10.1016/j.cell.2008.11.037

Dance, A., 2015. C9ORF72 Mice Point to Gain of Toxic Function in ALS, FTD [WWW Document]. Alzforum. URL http://www.alzforum.org/news/conference-coverage/c9orf72-mice-point-gain-toxic-function-als-ftd (accessed 12.8.15).

Daoud, H., Zhou, S., Noreau, A., Sabbagh, M., Belzil, V., Dionne-Laporte, A., Tranchant, C., Dion, P., Rouleau, G.A., 2012. Exome sequencing reveals SPG11 mutations causing juvenile ALS. Neurobiol. Aging 33, 839.e5–9. doi:10.1016/j.neurobiolaging.2011.11.012

Davidson, Y.S., Barker, H., Robinson, A.C., Thompson, J.C., Harris, J., Troakes, C., Smith, B., Al-Saraj, S., Shaw, C., Rollinson, S., Masuda-Suzukake, M., Hasegawa, M., Pickering-Brown, S., Snowden, J.S., Mann, D.M., 2014. Brain distribution of dipeptide repeat proteins in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. Acta Neuropathol. Commun. 2, 70. doi:10.1186/2051-5960-2-70

de Haro, M., Al-Ramahi, I., De Gouyon, B., Ukani, L., Rosa, A., Faustino, N.A., Ashizawa, T., Cooper, T.A., Botas, J., 2006. MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1. Hum. Mol. Genet. 15, 2138–2145. doi:10.1093/hmg/ddl137

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.-Y.R., Karydas, A., Seeley, W.W., Josephs, K.A., Coppola, G., Geschwind, D.H., Wszolek, Z.K., Feldman, H., Knopman, D.S., Petersen, R.C., Miller, B.L., Dickson, D.W., Boylan, K.B., Graff-Radford, N.R., Rademakers, R., 2011. Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. Neuron 72, 1–12. doi:10.1016/j.neuron.2011.09.011

Delatycki, M.B., Williamson, R., Forrest, S.M., 2000. Friedreich ataxia: an overview. J. Med. Genet. 37, 1–8.

Deng, H., Gao, K., Jankovic, J., 2014. The role of FUS gene variants in neurodegenerative diseases. Nat. Rev. Neurol. 10, 337–48. doi:10.1038/nrneurol.2014.78

Deng, H.-X., Chen, W., Hong, S.-T., Boycott, K.M., Gorrie, G.H., Siddique, N., Yang, Y., Fecto, F., Shi, Y., Zhai, H., Jiang, H., Hirano, M., Rampersaud, E., Jansen, G.H., Donkervoort, S., Bigio, E.H., Brooks, B.R., Ajroud, K., Sufit, R.L., Haines, J.L., Mugnaini, E., Pericak-Vance, M.A., Siddique, T., 2011. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature 477, 211–5. doi:10.1038/nature10353

Dias, N., Stein, C.A., 2002. Antisense Oligonucleotides: Basic Concepts and Mechanisms. Mol. Cancer Ther. 1, 347–355.

Dickmanns, A., Kehlenbach, R.H., Fahrenkrog, B., 2015. Nuclear Pore Complexes and Nucleocytoplasmic Transport: From Structure to Function to Disease. Int. Rev. Cell Mol. Biol. 320, 171–233. doi:10.1016/bs.ircmb.2015.07.010

Dobson-Stone, C., Hallupp, M., Loy, C.T., Thompson, E.M., Haan, E., Sue, C.M., Panegyres, P.K., Razquin, C., Seijo-Martínez, M., Rene, R., Gascon, J., Campdelacreu, J., Schmoll, B., Volk, A.E., Brooks, W.S., Schofield, P.R., Pastor, P., Kwok, J.B.J., 2013. C9ORF72 repeat expansion in Australian and Spanish frontotemporal dementia patients. PLoS One 8, e56899. doi:10.1371/journal.pone.0056899

Doi, H., Okamura, K., Bauer, P.O., Furukawa, Y., Shimizu, H., Kurosawa, M., Machida, Y., Miyazaki, H., Mitsui, K., Kuroiwa, Y., Nukina, N., 2008. RNA-binding protein TLS is a major nuclear aggregate-interacting protein in huntingtin exon 1 with expanded polyglutamine-expressing cells. J. Biol. Chem. 283, 6489–500. doi:10.1074/jbc.M705306200

Donnelly, C.J., Zhang, P.-W., Pham, J.T., Heusler, A.R., Mistry, N.A., Vidensky, S., Daley, E.L., Poth, E.M., Hoover, B., Fines, D.M., Maragakis, N., Tienari, P.J., Petrucelli, L., Traynor, B.J., Wang, J., Rigo, F., Bennett, C.F., Blackshaw, S., Sattler, R., Rothstein, J.D., 2013. RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. Neuron 80, 415–28. doi:10.1016/j.neuron.2013.10.015

Edbauer, D., Haass, C., 2015. An amyloid-like cascade hypothesis for C9orf72 ALS/FTD. Curr. Opin. Neurobiol. 36, 99–106. doi:10.1016/j.conb.2015.10.009

Evans-Galea, M. V, Carrodus, N., Rowley, S.M., Corben, L.A., Tai, G., Saffery, R., Galati, J.C., Wong, N.C., Craig, J.M., Lynch, D.R., Regner, S.R., Brocht, A.F.D., Perlman, S.L., Bushara, K.O., Gomez, C.M., Wilmot, G.R., Li, L., Varley, E., Delatycki, M.B., Sarsero, J.P., 2012. FXN methylation predicts expression and clinical outcome in Friedreich ataxia. Ann. Neurol. 71, 487–97. doi:10.1002/ana.22671

Farg, M.A., Sundaramoorthy, V., Sultana, J.M., Yang, S., Atkinson, R.A.K., Levina, V., Halloran, M.A., Gleeson, P.A., Blair, I.P., Soo, K.Y., King, A.E., Atkin, J.D., 2014. C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. Hum. Mol. Genet. 23, 3579–95.

doi:10.1093/hmg/ddu068

Fecto, F., 2011. <emph type="ital">SQSTM1</emph> Mutations in Familial and Sporadic Amyotrophic Lateral Sclerosis. Arch. Neurol. 68, 1440. doi:10.1001/archneurol.2011.250

Feiguin, F., Godena, V.K., Romano, G., D'Ambrogio, A., Klima, R., Baralle, F.E., 2009. Depletion of TDP-43 affects Drosophila motoneurons terminal synapsis and locomotive behavior. FEBS Lett. 583, 1586–92. doi:10.1016/j.febslet.2009.04.019

Fratta, P., Mizielinska, S., Nicoll, A.J., Zloh, M., Fisher, E.M.C., Parkinson, G., Isaacs, A.M., 2012. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. Sci. Rep. 2, 1016. doi:10.1038/srep01016

Fratta, P., Polke, J.M., Newcombe, J., Mizielinska, S., Lashley, T., Poulter, M., Beck, J., Preza, E., Devoy, A., Sidle, K., Howard, R., Malaspina, A., Orrell, R.W., Clarke, J., Lu, C.-H., Mok, K., Collins, T., Shoaii, M., Nanji, T., Wray, S., Adamson, G., Pittman, A., Renton, A.E., Traynor, B.J., Sweeney, M.G., Revesz, T., Houlden, H., Mead, S., Isaacs, A.M., Fisher, E.M.C., 2015. Screening a UK amyotrophic lateral sclerosis cohort provides evidence of multiple origins of the C9orf72 expansion. Neurobiol. Aging 36, 546.e1–546.e7. doi:10.1016/j.neurobiolaging.2014.07.037

Fratta, P., Poulter, M., Lashley, T., Rohrer, J.D., Polke, J.M., Beck, J., Ryan, N., Hensman, D., Mizielinska, S., Waite, A.J., Lai, M.-C., Gendron, T.F., Petrucelli, L., Fisher, E.M.C., Revesz, T., Warren, J.D., Collinge, J., Isaacs, A.M., Mead, S., 2013. Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. Acta Neuropathol. 126, 401–9. doi:10.1007/s00401-013-1147-0

Frehlick, L.J., Eirín-López, J.M., Ausió, J., 2007. New insights into the nucleophosmin/nucleoplasmin family of nuclear chaperones. Bioessays 29, 49–59. doi:10.1002/bies.20512

Freibaum, B.D., Lu, Y., Lopez-Gonzalez, R., Kim, N.C., Almeida, S., Lee, K.-H., Badders, N., Valentine, M., Miller, B.L., Wong, P.C., Petrucelli, L., Kim, H.J., Gao, F.-B., Taylor, J.P., 2015. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. Nature advance on. doi:10.1038/nature14974

Fu, Y., Pizzuti, A., Fenwick, R., King, J., Rajnarayan, S., Dunne, P., Dubel, J., Nasser, G., Ashizawa, T., de Jong, P., et, al., 1992. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science (80-.). 255, 1256–1258. doi:10.1126/science.1546326

García Moreno, L.., Cimadevilla, J.., González Pardo, H., Zahonero, M.., Arias, J.., 1997. NOR activity in hippocampal areas during the postnatal development and ageing. Mech. Ageing Dev. 97, 173–181. doi:10.1016/S0047-6374(97)00054-7

Gass, J., Cannon, A., Mackenzie, I.R., Boeve, B., Baker, M., Adamson, J., Crook, R., Melquist, S., Kuntz, K., Petersen, R., Josephs, K., Pickering-Brown, S.M., Graff-Radford, N., Uitti,

R., Dickson, D., Wszolek, Z., Gonzalez, J., Beach, T.G., Bigio, E., Johnson, N., Weintraub, S., Mesulam, M., White, C.L., Woodruff, B., Caselli, R., Hsiung, G.-Y., Feldman, H., Knopman, D., Hutton, M., Rademakers, R., 2006. Mutations in progranulin are a major cause of ubiquitin-positive frontotemporal lobar degeneration. Hum. Mol. Genet. 15, 2988–3001. doi:10.1093/hmg/ddl241

Gendron, T.F., Bieniek, K.F., Zhang, Y.-J., Jansen-West, K., Ash, P.E.A., Caulfield, T., Daughrity, L., Dunmore, J.H., Castanedes-Casey, M., Chew, J., Cosio, D.M., van Blitterswijk, M., Lee, W.C., Rademakers, R., Boylan, K.B., Dickson, D.W., Petrucelli, L., 2013. Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol. doi:10.1007/s00401-013-1192-8

Gendron, T.F., van Blitterswijk, M., Bieniek, K.F., Daughrity, L.M., Jiang, J., Rush, B.K., Pedraza, O., Lucas, J.A., Murray, M.E., Desaro, P., Robertson, A., Overstreet, K., Thomas, C.S., Crook, J.E., Castanedes-Casey, M., Rousseau, L., Josephs, K.A., Parisi, J.E., Knopman, D.S., Petersen, R.C., Boeve, B.F., Graff-Radford, N.R., Rademakers, R., Lagier-Tourenne, C., Edbauer, D., Cleveland, D.W., Dickson, D.W., Petrucelli, L., Boylan, K.B., 2015. Cerebellar c9RAN proteins associate with clinical and neuropathological characteristics of C9ORF72 repeat expansion carriers. Acta Neuropathol. doi:10.1007/s00401-015-1474-4

Gijselinck, I., Engelborghs, S., Maes, G., Cuijt, I., Peeters, K., Mattheijssens, M., Joris, G., Cras, P., Martin, J.-J., De Deyn, P.P., Kumar-Singh, S., Van Broeckhoven, C., Cruts, M., 2010. Identification of 2 Loci at chromosomes 9 and 14 in a multiplex family with frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Arch. Neurol. 67, 606–16. doi:10.1001/archneurol.2010.82

Gijselinck, I., Van Broeckhoven, C., Cruts, M., 2008. Granulin mutations associated with frontotemporal lobar degeneration and related disorders: an update. Hum. Mutat. 29, 1373–86. doi:10.1002/humu.20785

Gijselinck, I., Van Langenhove, T., van der Zee, J., Sleegers, K., Philtjens, S., Kleinberger, G., Janssens, J., Bettens, K., Van Cauwenberghe, C., Pereson, S., Engelborghs, S., Sieben, A., De Jonghe, P., Vandenberghe, R., Santens, P., De Bleecker, J., Maes, G., Bäumer, V., Dillen, L., Joris, G., Cuijt, I., Corsmit, E., Elinck, E., Van Dongen, J., Vermeulen, S., Van den Broeck, M., Vaerenberg, C., Mattheijssens, M., Peeters, K., Robberecht, W., Cras, P., Martin, J.J., De Deyn, P.P., Cruts, M., Van Broeckhoven, C., 2012. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: A gene identification study. Lancet Neurol. 11, 54–65. doi:10.1016/S1474-4422(11)70261-7

Gijselinck, I., Van Mossevelde, S., van der Zee, J., Sieben, A., Engelborghs, S., De Bleecker, J., Ivanoiu, A., Deryck, O., Edbauer, D., Zhang, M., Heeman, B., Bäumer, V., Van den Broeck, M., Mattheijssens, M., Peeters, K., Rogaeva, E., De Jonghe, P., Cras, P., Martin, J.-J., de Deyn, P.P., Cruts, M., Van Broeckhoven, C., 2015. The C9orf72 repeat size

correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol. Psychiatry. doi:10.1038/mp.2015.159

Gomez-Deza, J., Lee, Y., Troakes, C., Nolan, M., Al-Sarraj, S., Gallo, J.-M., Shaw, C.E., 2015. Dipeptide repeat protein inclusions are rare in the spinal cord and almost absent from motor neurons in C9ORF72 mutant amyotrophic lateral sclerosis and are unlikely to cause their degeneration. Acta Neuropathol. Commun. 3, 38. doi:10.1186/s40478-015-0218-y

Gómez-Tortosa, E., Gallego, J., Guerrero-López, R., Marcos, A., Gil-Neciga, E., Sainz, M.J., Díaz, A., Franco-Macías, E., Trujillo-Tiebas, M.J., Ayuso, C., Pérez-Pérez, J., 2013. C9ORF72 hexanucleotide expansions of 20-22 repeats are associated with frontotemporal deterioration. Neurology 80, 366–70. doi:10.1212/WNL.0b013e31827f08ea

Greene, E., Mahishi, L., Entezam, A., Kumari, D., Usdin, K., 2007. Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. Nucleic Acids Res. 35, 3383–90. doi:10.1093/nar/gkm271

Greenway, M.J., Andersen, P.M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., Patterson, V., Swingler, R., Kieran, D., Prehn, J., Morrison, K.E., Green, A., Acharya, K.R., Brown, R.H., Hardiman, O., 2006. ANG mutations segregate with familial and “sporadic” amyotrophic lateral sclerosis. Nat. Genet. 38, 411–3. doi:10.1038/ng1742

Grummt, I., 2013. The nucleolus—guardian of cellular homeostasis and genome integrity. Chromosoma 122, 487–97. doi:10.1007/s00412-013-0430-0

Hadano, S., Hand, C.K., Osuga, H., Yanagisawa, Y., Otomo, A., Devon, R.S., Miyamoto, N., Showguchi-Miyata, J., Okada, Y., Singaraja, R., Figlewicz, D.A., Kwiatkowski, T., Hosler, B.A., Sagie, T., Skaug, J., Nasir, J., Brown, R.H., Scherer, S.W., Rouleau, G.A., Hayden, M.R., Ikeda, J.E., 2001. A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. Nat. Genet. 29, 166–73. doi:10.1038/ng1001-166

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A.J., Shaw, P.G., Kim, M.-S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., Rothstein, J.D., Wang, J., 2014. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature 507, 195–200. doi:10.1038/nature13124

Hariharan, N., Sussman, M.A., 2014. Stressing on the nucleolus in cardiovascular disease. Biochim. Biophys. Acta 1842, 798–801. doi:10.1016/j.bbadis.2013.09.016

Harms, M.B., Cady, J., Zaidman, C., Cooper, P., Bali, T., Allred, P., Cruchaga, C., Baughn, M., Libby, R.T., Pestronk, A., Goate, A., Ravits, J., Baloh, R.H., 2013. Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. Neurobiol. Aging 34, 2234.e13–9. doi:10.1016/j.neurobiolaging.2013.03.006

Hetman, M., Pietrzak, M., 2012. Emerging roles of the neuronal nucleolus. Trends Neurosci.

35, 305–14. doi:10.1016/j.tins.2012.01.002

Hideyama, T., Yamashita, T., Aizawa, H., Tsuji, S., Kakita, A., Takahashi, H., Kwak, S., 2012. Profound downregulation of the RNA editing enzyme ADAR2 in ALS spinal motor neurons. Neurobiol. Dis. 45, 1121–8. doi:10.1016/j.nbd.2011.12.033

Holm, I.E., Englund, E., Mackenzie, I.R.A., Johannsen, P., Isaacs, A.M., 2007. A reassessment of the neuropathology of frontotemporal dementia linked to chromosome 3. J. Neuropathol. Exp. Neurol. 66, 884–91. doi:10.1097/nen.0b013e3181567f02

Holm, I.E., Isaacs, A.M., Mackenzie, I.R.A., 2009. Absence of FUS-immunoreactive pathology in frontotemporal dementia linked to chromosome 3 (FTD-3) caused by mutation in the CHMP2B gene. Acta Neuropathol. 118, 719–20. doi:10.1007/s00401-009-0593-1

Hsiung, G.-Y.R., DeJesus-Hernandez, M., Feldman, H.H., Sengdy, P., Bouchard-Kerr, P., Dwosh, E., Butler, R., Leung, B., Fok, A., Rutherford, N.J., Baker, M., Rademakers, R., Mackenzie, I.R.A., 2012. Clinical and pathological features of familial frontotemporal dementia caused by C9ORF72 mutation on chromosome 9p. Brain 135, 709–22. doi:10.1093/brain/awr354

Hukema, R.K., Riemslagh, F.W., Melhem, S., van der Linde, H.C., Severijnen, L.-A., Edbauer, D., Maas, A., Charlet-Berguerand, N., Willemsen, R., van Swieten, J.C., 2014. A new inducible transgenic mouse model for C9orf72-associated GGGGCC repeat expansion supports a gain-of-function mechanism in C9orf72 associated ALS and FTD. Acta Neuropathol. Commun. 2, 166. doi:10.1186/s40478-014-0166-y

Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A., Hackett, J., Adamson, J., Lincoln, S., Dickson, D., Davies, P., Petersen, R.C., Stevens, M., de Graaff, E., Wauters, E., van Baren, J., Hillebrand, M., Joosse, M., Kwon, J.M., Nowotny, P., Che, L.K., Norton, J., Morris, J.C., Reed, L.A., Trojanowski, J., Basun, H., Lannfelt, L., Neystat, M., Fahn, S., Dark, F., Tannenberg, T., Dodd, P.R., Hayward, N., Kwok, J.B., Schofield, P.R., Andreadis, A., Snowden, J., Craufurd, D., Neary, D., Owen, F., Oostra, B.A., Hardy, J., Goate, A., van Swieten, J., Mann, D., Lynch, T., Heutink, P., 1998. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature 393, 702–5. doi:10.1038/31508

Iacono, D., Markesbery, W.R., Gross, M., Pletnikova, O., Rudow, G., Zandi, P., Troncoso, J.C., 2009. The Nun Study: Clinically silent AD, neuronal hypertrophy, and linguistic skills in early life. Neurology 73, 665–673. doi:10.1212/WNL.0b013e3181b01077

Iguchi, Y., Katsuno, M., Niwa, J., Yamada, S., Sone, J., Waza, M., Adachi, H., Tanaka, F., Nagata, K., Arimura, N., Watanabe, T., Kaibuchi, K., Sobue, G., 2009. TDP-43 depletion induces neuronal cell damage through dysregulation of Rho family GTPases. J. Biol. Chem. 284, 22059–66. doi:10.1074/jbc.M109.012195

Ishiura, H., Takahashi, Y., Mitsui, J., Yoshida, S., Kihira, T., Kokubo, Y., Kuzuhara, S., Ranum, L.P.W., Tamaoki, T., Ichikawa, Y., Date, H., Goto, J., Tsuji, S., 2012. C9ORF72 repeat

expansion in amyotrophic lateral sclerosis in the Kii peninsula of Japan. Arch. Neurol. 69, 1154–8. doi:10.1001/archneurol.2012.1219

Jin, P., Duan, R., Qurashi, A., Qin, Y., Tian, D., Rosser, T.C., Liu, H., Feng, Y., Warren, S.T., 2007. Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. Neuron 55, 556–64. doi:10.1016/j.neuron.2007.07.020

Johnson, J.O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V.M., Trojanowski, J.Q., Gibbs, J.R., Brunetti, M., Gronka, S., Wuu, J., Ding, J., McCluskey, L., Martinez-Lage, M., Falcone, D., Hernandez, D.G., Arepalli, S., Chong, S., Schymick, J.C., Rothstein, J., Landi, F., Wang, Y.-D., Calvo, A., Mora, G., Sabatelli, M., Monsurrò, M.R., Battistini, S., Salvi, F., Spataro, R., Sola, P., Borghero, G., Galassi, G., Scholz, S.W., Taylor, J.P., Restagno, G., Chiò, A., Traynor, B.J., 2010. Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron 68, 857–64. doi:10.1016/j.neuron.2010.11.036

Jovičić, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S.B., Paul, J.W., Sun, S., Herdy, J.R., Bieri, G., Kramer, N.J., Gage, F.H., Van Den Bosch, L., Robberecht, W., Gitler, A.D., 2015. Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. Nat. Neurosci. 18, 1226–1229. doi:10.1038/nn.4085

Kalita, K., Makonchuk, D., Gomes, C., Zheng, J.-J., Hetman, M., 2008. Inhibition of nucleolar transcription as a trigger for neuronal apoptosis. J. Neurochem. 105, 2286–99. doi:10.1111/j.1471-4159.2008.05316.x

Kato, S., Takikawa, M., Nakashima, K., Hirano, A., Cleveland, D.W., Kusaka, H., Shibata, N., Kato, M., Nakano, I., Ohama, E., 2000. New consensus research on neuropathological aspects of familial amyotrophic lateral sclerosis with superoxide dismutase 1 (SOD1) gene mutations: inclusions containing SOD1 in neurons and astrocytes. Amyotroph. Lateral Scler. Other Motor Neuron Disord. 1, 163–84.

Kertesz, A., Munoz, D., 2004. Relationship between frontotemporal dementia and corticobasal degeneration/progressive supranuclear palsy. Dement. Geriatr. Cogn. Disord. 17, 282–6. doi:10.1159/000077155

Khlistunova, I., Biernat, J., Wang, Y., Pickhardt, M., von Bergen, M., Gazova, Z., Mandelkow, E., Mandelkow, E.-M., 2006. Inducible expression of Tau repeat domain in cell models of tauopathy: aggregation is toxic to cells but can be reversed by inhibitor drugs. J. Biol. Chem. 281, 1205–14. doi:10.1074/jbc.M507753200

Kiernan, M.C., Vucic, S., Cheah, B.C., Turner, M.R., Eisen, A., Hardiman, O., Burrell, J.R., Zoing, M.C., 2011. Amyotrophic lateral sclerosis. Lancet 377, 942–55. doi:10.1016/S0140-6736(10)61156-7

Kim, H.J., Kim, N.C., Wang, Y.-D., Scarborough, E.A., Moore, J., Diaz, Z., MacLea, K.S., Freibaum, B., Li, S., Molliex, A., Kanagaraj, A.P., Carter, R., Boylan, K.B., Wojtas, A.M., Rademakers, R., Pinkus, J.L., Greenberg, S.A., Trojanowski, J.Q., Traynor, B.J., Smith,

B.N., Topp, S., Gkazi, A.-S., Miller, J., Shaw, C.E., Kottlors, M., Kirschner, J., Pestronk, A., Li, Y.R., Ford, A.F., Gitler, A.D., Benatar, M., King, O.D., Kimonis, V.E., Ross, E.D., Weihl, C.C., Shorter, J., Taylor, J.P., 2013. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature 495, 467–73. doi:10.1038/nature11922

Klesert, T.R., Otten, A.D., Bird, T.D., Tapscott, S.J., 1997. Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. Nat. Genet. 16, 402–6. doi:10.1038/ng0897-402

Koch, K.S., Leffert, H.L., 1998. Giant hairpins formed by CUG repeats in myotonic dystrophy messenger RNAs might sterically block RNA export through nuclear pores. J. Theor. Biol. 192, 505–14. doi:10.1006/jtbi.1998.0679

Koppers, M., Blokhuis, A.M., Westeneng, H.-J., Terpstra, M.L., Zundel, C.A.C., Vieira de Sá, R., Schellevis, R.D., Waite, A.J., Blake, D.J., Veldink, J.H., van den Berg, L.H., Pasterkamp, R.J., 2015. C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. Ann. Neurol. 78, 426–438. doi:10.1002/ana.24453

Korade-Mirnics, Z., 1999. Myotonic dystrophy: tissue-specific effect of somatic CTG expansions on allele-specific DMAHP/SIX5 expression. Hum. Mol. Genet. 8, 1017–1023. doi:10.1093/hmg/8.6.1017

Kovalevich, J., Langford, D., 2013. Considerations for the use of SH-SY5Y neuroblastoma cells in neurobiology. Methods Mol. Biol. 1078, 9–21. doi:10.1007/978-1-62703-640-5_2

Kozak, M., 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc. Natl. Acad. Sci. 87, 8301–8305. doi:10.1073/pnas.87.21.8301

Kozak, M., 1989. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. Mol. Cell. Biol. 9, 5073–5080. doi:10.1128/MCB.9.11.5073

Kraemer, B.C., Schuck, T., Wheeler, J.M., Robinson, L.C., Trojanowski, J.Q., Lee, V.M.Y., Schellenberg, G.D., 2010. Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. Acta Neuropathol. 119, 409–19. doi:10.1007/s00401-010-0659-0

Krecic, A.M., Swanson, M.S., 1999. hnRNP complexes: composition, structure, and function. Curr. Opin. Cell Biol. 11, 363–71. doi:10.1016/S0955-0674(99)80051-9

Kwiatkowski, T.J., Bosco, D.A., Leclerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., Valdmanis, P., Rouleau, G.A., Hosler, B.A., Cortelli, P., de Jong, P.J., Yoshinaga, Y., Haines, J.L., Pericak-Vance, M.A., Yan, J., Ticozzi, N., Siddique, T., McKenna-Yasek, D., Sapp, P.C., Horvitz, H.R., Landers, J.E., Brown, R.H., 2009. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science 323, 1205–8. doi:10.1126/science.1166066

Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., McKnight, S.L., 2014. Poly-dipeptides encoded by the C9ORF72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science 1139. doi:10.1126/science.1254917

Laaksovirta, H., Peuralinna, T., Schymick, J.C., Scholz, S.W., Lai, S.-L., Myllykangas, L., Sulkava, R., Jansson, L., Hernandez, D.G., Gibbs, J.R., Nalls, M.A., Heckerman, D., Tienari, P.J., Traynor, B.J., 2010. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. Lancet Neurol. 9, 978–85. doi:10.1016/S1474-4422(10)70184-8

Lagier-Tourenne, C., Baughn, M., Rigo, F., Sun, S., Liu, P., Li, H.-R., Jiang, J., Watt, A.T., Chun, S., Katz, M., Qiu, J., Sun, Y., Ling, S.-C., Zhu, Q., Polymenidou, M., Drenner, K., Artates, J.W., McAlonis-Downes, M., Markmiller, S., Hutt, K.R., Pizzo, D.P., Cady, J., Harms, M.B., Baloh, R.H., Vandenberg, S.R., Yeo, G.W., Fu, X.-D., Bennett, C.F., Cleveland, D.W., Ravits, J., 2013. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc. Natl. Acad. Sci. U. S. A. 110, E4530–9. doi:10.1073/pnas.1318835110

Lagier-Tourenne, C., Polymenidou, M., Cleveland, D.W., 2010. TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Hum. Mol. Genet. 19, R46–R64.

Langbehn, D.R., Brinkman, R.R., Falush, D., Paulsen, J.S., Hayden, M.R., 2004. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clin. Genet. 65, 267–77. doi:10.1111/j.1399-0004.2004.00241.x

Lattante, S., Rouleau, G.A., Kabashi, E., 2013. TARDBP and FUS mutations associated with amyotrophic lateral sclerosis: summary and update. Hum. Mutat. 34, 812–26. doi:10.1002/humu.22319

Le Ber, I., Camuzat, A., Berger, E., Hannequin, D., Laquerrière, A., Golfier, V., Seilhean, D., Viennet, G., Couratier, P., Verpillat, P., Heath, S., Camu, W., Martinaud, O., Lacomblez, L., Vercelletto, M., Salachas, F., Sellal, F., Didic, M., Thomas-Anterion, C., Puel, M., Michel, B.-F., Besse, C., Duyckaerts, C., Meininger, V., Campion, D., Dubois, B., Brice, A., 2009. Chromosome 9p-linked families with frontotemporal dementia associated with motor neuron disease. Neurology 72, 1669–76. doi:10.1212/WNL.0b013e3181a55f1c

Le Ber, I., Camuzat, A., Guerreiro, R., Bouya-Ahmed, K., Bras, J., Nicolas, G., Gabelle, A., Didic, M., De Septenville, A., Millecamps, S., Lenglet, T., Latouche, M., Kabashi, E., Campion, D., Hannequin, D., Hardy, J., Brice, A., 2013. SQSTM1 mutations in French patients with frontotemporal dementia or frontotemporal dementia with amyotrophic lateral sclerosis. JAMA Neurol. 70, 1403–10. doi:10.1001/jamaneurol.2013.3849

Lee, J., Hwang, Y.J., Ryu, H., Kowall, N.W., Ryu, H., 2014. Nucleolar dysfunction in Huntington's disease. Biochim. Biophys. Acta 1842, 785–90. doi:10.1016/j.bbadis.2013.09.017

Lee, Y.B., Chen, H.J., Peres, J., Gomez-Deza, J., Attig, J., Štalekar, M., Troakes, C., Nishimura, A., Scotter, E., Vance, C., Adachi, Y., Sardone, V., Miller, J., Smith, B., Gallo, J.M., Ule, J., Hirth, F., Rogelj, B., Houart, C., Shaw, C., 2013. Hexanucleotide repeats in ALS/FTD form length-dependent RNA Foci, sequester RNA binding proteins, and are neurotoxic. Cell Rep. 5, 1178–1186. doi:10.1016/j.celrep.2013.10.049

Levine, T.P., Daniels, R.D., Gatta, A.T., Wong, L.H., Hayes, M.J., 2013. The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. Bioinformatics 29, 499–503. doi:10.1093/bioinformatics/bts725

Lewis, F., Karlsberg Schaffer, S., Sussex, J., O'Neil, P., Crockcroft, L., 2014. The Trajectory of Dementia in the UK – Making a Difference.

Li, Y., Ray, P., Rao, E.J., Shi, C., Guo, W., Chen, X., Woodruff, E.A., Fushimi, K., Wu, J.Y., 2010. A Drosophila model for TDP-43 proteinopathy. Proc. Natl. Acad. Sci. U. S. A. 107, 3169–74. doi:10.1073/pnas.0913602107

Lillo, P., Hodges, J.R., 2009. Frontotemporal dementia and motor neurone disease: overlapping clinic-pathological disorders. J. Clin. Neurosci. 16, 1131–5. doi:10.1016/j.jocn.2009.03.005

Liu, E.Y., Russ, J., Wu, K., Neal, D., Suh, E., McNally, A.G., Irwin, D.J., Van Deerlin, V.M., Lee, E.B., 2014. C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. Acta Neuropathol. 128, 525–541. doi:10.1007/s00401-014-1286-y

Logroscino, G., Traynor, B.J., Hardiman, O., Chiò, A., Mitchell, D., Swingler, R.J., Millul, A., Benn, E., Beghi, E., 2010. Incidence of amyotrophic lateral sclerosis in Europe. J. Neurol. Neurosurg. Psychiatry 81, 385–90. doi:10.1136/jnnp.2009.183525

Lomen-Hoerth, C., Anderson, T., Miller, B., 2002. The overlap of amyotrophic lateral sclerosis and frontotemporal dementia. Neurology 59, 1077–1079. doi:10.1212/WNL.59.7.1077

Lomen-Hoerth, C., Murphy, J., Langmore, S., Kramer, J.H., Olney, R.K., Miller, B., 2003. Are amyotrophic lateral sclerosis patients cognitively normal? Neurology 60, 1094–7.

Luty, A.A., Kwok, J.B.J., Dobson-Stone, C., Loy, C.T., Coupland, K.G., Karlström, H., Sobow, T., Tchorzewska, J., Maruszak, A., Barcikowska, M., Panegyres, P.K., Zekanowski, C., Brooks, W.S., Williams, K.L., Blair, I.P., Mather, K.A., Sachdev, P.S., Halliday, G.M., Schofield, P.R., 2010. Sigma nonopioid intracellular receptor 1 mutations cause frontotemporal lobar degeneration-motor neuron disease. Ann. Neurol. 68, 639–49. doi:10.1002/ana.22274

Luty, A.A., Kwok, J.B.J., Thompson, E.M., Blumbergs, P., Brooks, W.S., Loy, C.T., Dobson-Stone, C., Panegyres, P.K., Hecker, J., Nicholson, G.A., Halliday, G.M., Schofield, P.R., 2008. Pedigree with frontotemporal lobar degeneration--motor neuron disease and Tar DNA binding protein-43 positive neuropathology: genetic linkage to chromosome 9. BMC Neurol. 8, 32. doi:10.1186/1471-2377-8-32

Ma, A.S.W., Moran-Jones, K., Shan, J., Munro, T.P., Snee, M.J., Hoek, K.S., Smith, R., 2002. Heterogeneous Nuclear Ribonucleoprotein A3, a Novel RNA Trafficking Response Element-binding Protein. J. Biol. Chem. 277, 18010–18020. doi:10.1074/jbc.M200050200

Mackenzie, I.R.A., Arzberger, T., Kremmer, E., Troost, D., Lorenzl, S., Mori, K., Weng, S.-M., Haass, C., Kretzschmar, H.A., Edbauer, D., Neumann, M., 2013. Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. Acta Neuropathol. 126, 859–79. doi:10.1007/s00401-013-1181-y

Mackenzie, I.R.A., Bigio, E.H., Ince, P.G., Geser, F., Neumann, M., Cairns, N.J., Kwong, L.K., Forman, M.S., Ravits, J., Stewart, H., Eisen, A., McClusky, L., Kretzschmar, H.A., Monoranu, C.M., Highley, J.R., Kirby, J., Siddique, T., Shaw, P.J., Lee, V.M.-Y., Trojanowski, J.Q., 2007. Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. Ann. Neurol. 61, 427–34. doi:10.1002/ana.21147

Mackenzie, I.R.A., Frick, P., Grässer, F.A., Gendron, T.F., Petrucelli, L., Cashman, N.R., Edbauer, D., Kremmer, E., Prudlo, J., Troost, D., Neumann, M., 2015. Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. Acta Neuropathol. doi:10.1007/s00401-015-1476-2

Mackenzie, I.R.A., Frick, P., Neumann, M., 2014. The neuropathology associated with repeat expansions in the C9ORF72 gene. Acta Neuropathol. 127, 347–57. doi:10.1007/s00401-013-1232-4

Mackenzie, I.R.A., Neumann, M., Baborie, A., Sampathu, D.M., Du Plessis, D., Jaros, E., Perry, R.H., Trojanowski, J.Q., Mann, D.M.A., Lee, V.M.Y., 2011. A harmonized classification system for FTLD-TDP pathology. Acta Neuropathol. 122, 111–3. doi:10.1007/s00401-011-0845-8

Mackenzie, I.R.A., Neumann, M., Bigio, E.H., Cairns, N.J., Alafuzoff, I., Kril, J., Kovacs, G.G., Ghetti, B., Halliday, G., Holm, I.E., Ince, P.G., Kamphorst, W., Revesz, T., Rozemuller, A.J.M., Kumar-Singh, S., Akiyama, H., Baborie, A., Spina, S., Dickson, D.W., Trojanowski, J.Q., Mann, D.M.A., 2010a. Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: an update. Acta Neuropathol. 119, 1–4. doi:10.1007/s00401-009-0612-2

Mackenzie, I.R.A., Rademakers, R., Neumann, M., 2010b. TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. Lancet. Neurol. 9, 995–1007. doi:10.1016/S1474-4422(10)70195-2

Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barceló, J., O'Hoy, K., 1992. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science 255, 1253–1255. doi:10.1126/science.1546325

Mahoney, C.J., Beck, J., Rohrer, J.D., Lashley, T., Mok, K., Shakespeare, T., Yeatman, T., Warrington, E.K., Schott, J.M., Fox, N.C., Rossor, M.N., Hardy, J., Collinge, J., Revesz, T., Mead, S., Warren, J.D., 2012. Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features. Brain 135, 736–50. doi:10.1093/brain/awr361

Majounie, E., Abramzon, Y., Renton, A.E., Perry, R., Bassett, S.S., Pletnikova, O., Troncoso, J.C., Hardy, J., Singleton, A.B., Traynor, B.J., 2012a. Repeat expansion in C9ORF72 in Alzheimer's disease. N. Engl. J. Med. 366, 283–4. doi:10.1056/NEJMc1113592

Majounie, E., Renton, A.E., Mok, K., Dopper, E.G.P., Waite, A., Rollinson, S., Chiò, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J., van Swieten, J.C., Abramzon, Y., Johnson, J.O., Sendtner, M., Pamphlett, R., Orrell, R.W., Mead, S., Sidle, K.C., Houlden, H., Rohrer, J.D., Morrison, K.E., Pall, H., Talbot, K., Ansorge, O., Hernandez, D.G., Arepalli, S., Sabatelli, M., Mora, G., Corbo, M., Giannini, F., Calvo, A., Englund, E., Borghero, G., Floris, G.L., Remes, A.M., Laaksovirta, H., McCluskey, L., Trojanowski, J.Q., Van Deerlin, V.M., Schellenberg, G.D., Nalls, M. a., Drory, V.E., Lu, C.S., Yeh, T.H., Ishiura, H., Takahashi, Y., Tsuji, S., Le Ber, I., Brice, A., Drepper, C., Williams, N., Kirby, J., Shaw, P., Hardy, J., Tienari, P.J., Heutink, P., Morris, H.R., Pickering-Brown, S., Traynor, B.J., 2012b. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. Lancet Neurol. 11, 323–330. doi:10.1016/S1474-4422(12)70043-1

Mankodi, A., 2001. Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. Hum. Mol. Genet. 10, 2165–2170. doi:10.1093/hmg/10.19.2165

Mankodi, A., Takahashi, M.P., Jiang, H., Beck, C.L., Bowers, W.J., Moxley, R.T., Cannon, S.C., Thornton, C.A., 2002. Expanded CUG Repeats Trigger Aberrant Splicing of ClC-1 Chloride Channel Pre-mRNA and Hyperexcitability of Skeletal Muscle in Myotonic Dystrophy. Mol. Cell 10, 35–44. doi:10.1016/S1097-2765(02)00563-4

Mann, D.M., Rollinson, S., Robinson, A., Bennion Callister, J., Thompson, J.C., Snowden, J.S., Gendron, T., Petrucelli, L., Masuda-Suzukake, M., Hasegawa, M., Davidson, Y., Pickering-Brown, S., 2013. Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. Acta Neuropathol. Commun. 1, 68. doi:10.1186/2051-5960-1-68

Mann, D.M.A., Neary, D., Yates, P.O., Lincoln, J., Snowden, J.S., Stanworth, P., 1981. Neurofibrillary Pathology And Protein Synthetic Capability In Nerve Cells In Alzheimer's Disease. Neuropathol. Appl. Neurobiol. 7, 37–47. doi:10.1111/j.1365-2990.1981.tb00230.x

Mann, D.M.A., Yates, P.O., 1982. Pathogenesis of Parkinson's disease. Arch. Neurol. 39, 545–549.

Martinez-Vicente, M., Sovak, G., Cuervo, A.M., 2005. Protein degradation and aging. Exp.

Gerontol. 40, 622–33. doi:10.1016/j.exger.2005.07.005

Maruyama, H., Morino, H., Ito, H., Izumi, Y., Kato, H., Watanabe, Y., Kinoshita, Y., Kamada, M., Nodera, H., Suzuki, H., Komure, O., Matsuura, S., Kobatake, K., Morimoto, N., Abe, K., Suzuki, N., Aoki, M., Kawata, A., Hirai, T., Kato, T., Ogasawara, K., Hirano, A., Takumi, T., Kusaka, H., Hagiwara, K., Kaji, R., Kawakami, H., 2010. Mutations of optineurin in amyotrophic lateral sclerosis. Nature 465, 223–6. doi:10.1038/nature08971

May, S., Hornburg, D., Schludi, M.H., Arzberger, T., Rentzsch, K., Schwenk, B.M., Grässer, F.A., Mori, K., Kremmer, E., Banzhaf-Strathmann, J., Mann, M., Meissner, F., Edbauer, D., 2014. C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. Acta Neuropathol. doi:10.1007/s00401-014-1329-4

McGoldrick, P., Joyce, P.I., Fisher, E.M.C., Greensmith, L., 2013. Rodent models of amyotrophic lateral sclerosis. Biochim. Biophys. Acta 1832, 1421–36. doi:10.1016/j.bbadis.2013.03.012

McMillan, C.T., Russ, J., Wood, E.M., Irwin, D.J., Grossman, M., McCluskey, L., Elman, L., Van Deerlin, V., Lee, E.B., 2015. C9orf72 promoter hypermethylation is neuroprotective: Neuroimaging and neuropathologic evidence. Neurology 84, 1622–30. doi:10.1212/WNL.0000000000001495

McMurray, C.T., 2010. Mechanisms of trinucleotide repeat instability during human development. Nat. Rev. Genet. 11, 786–99. doi:10.1038/nrg2828

Meyer, D.E., Chilkoti, A., 2002. Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: examples from the elastin-like polypeptide system. Biomacromolecules 3, 357–367.

Millecamps, S., Boillée, S., Le Ber, I., Seilhean, D., Teyssou, E., Giraudeau, M., Moigneu, C., Vandenberghe, N., Danel-Brunaud, V., Corcia, P., Pradat, P.-F., Le Forestier, N., Lacomblez, L., Bruneteau, G., Camu, W., Brice, A., Cazeneuve, C., Leguern, E., Meininger, V., Salachas, F., 2012. Phenotype difference between ALS patients with expanded repeats in C9ORF72 and patients with mutations in other ALS-related genes. J. Med. Genet. 49, 258–63. doi:10.1136/jmedgenet-2011-100699

Miller, J.W., Urbinati, C.R., Teng-Umnuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A., Swanson, M.S., 2000. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. EMBO J. 19, 4439–48. doi:10.1093/emboj/19.17.4439

Miller, R.G., Mitchell, J.D., Lyon, M., Moore, D.H., 2007. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). Cochrane database Syst. Rev. CD001447. doi:10.1002/14651858.CD001447.pub2

Millevoi, S., Moine, H., Vagner, S., 2012. G-quadruplexes in RNA biology. Wiley Interdiscip.

Rev. RNA 3, 495–507. doi:10.1002/wrna.1113

Mitchell, J., Paul, P., Chen, H.-J., Morris, A., Payling, M., Falchi, M., Habgood, J., Panoutsou, S., Winkler, S., Tisato, V., Hajitou, A., Smith, B., Vance, C., Shaw, C., Mazarakis, N.D., de Belleroche, J., 2010. Familial amyotrophic lateral sclerosis is associated with a mutation in D-amino acid oxidase. Proc. Natl. Acad. Sci. U. S. A. 107, 7556–61. doi:10.1073/pnas.0914128107

Mizielinska, S., Grönke, S., Niccoli, T., Ridler, C.E., Clayton, E.L., Devoy, A., Moens, T., Norona, F.E., Woollacott, I.O.C., Pietrzyk, J., Cleverley, K., Nicoll, A.J., Pickering-Brown, S., Dols, J., Cabecinha, M., Hendrich, O., Fratta, P., Fisher, E.M.C., Partridge, L., Isaacs, A.M., 2014. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science 1192. doi:10.1126/science.1256800

Mizielinska, S., Lashley, T., Norona, F.E., Clayton, E.L., Ridler, C.E., Fratta, P., Isaacs, A.M., 2013. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. Acta Neuropathol. 126, 845–57. doi:10.1007/s00401-013-1200-z

Mongelard, F., Bouvet, P., 2007. Nucleolin: a multiFACeTed protein. Trends Cell Biol. 17, 80–6. doi:10.1016/j.tcb.2006.11.010

Montanaro, L., Treré, D., Derenzini, M., 2008. Nucleolus, ribosomes, and cancer. Am. J. Pathol. 173, 301–10. doi:10.2353/ajpath.2008.070752

Morales, F., Couto, J.M., Higham, C.F., Hogg, G., Cuenca, P., Braida, C., Wilson, R.H., Adam, B., del Valle, G., Brian, R., Sittenfeld, M., Ashizawa, T., Wilcox, A., Wilcox, D.E., Monckton, D.G., 2012. Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. Hum. Mol. Genet. 21, 3558–67. doi:10.1093/hmg/dds185

Mori, K., Arzberger, T., Grässer, F.A., Gijselinck, I., May, S., Rentzsch, K., Weng, S.-M., Schludi, M.H., van der Zee, J., Cruts, M., Van Broeckhoven, C., Kremmer, E., Kretzschmar, H.A., Haass, C., Edbauer, D., 2013a. Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. Acta Neuropathol. doi:10.1007/s00401-013-1189-3

Mori, K., Lammich, S., Mackenzie, I.R. a, Forné, I., Zilow, S., Kretzschmar, H., Edbauer, D., Janssens, J., Kleinberger, G., Cruts, M., Herms, J., Neumann, M., Van Broeckhoven, C., Arzberger, T., Haass, C., 2013b. hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. Acta Neuropathol. 125, 413–23. doi:10.1007/s00401-013-1088-7

Mori, K., Weng, S.-M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H. a, Cruts, M., Van Broeckhoven, C., Haass, C., Edbauer, D., 2013c. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science 339, 1335–8. doi:10.1126/science.1232927

Morita, M., Al-Chalabi, A., Andersen, P.M., Hosler, B., Sapp, P., Englund, E., Mitchell, J.E., Habgood, J.J., De Belleroche, J., Xi, J., Jongjaroenprasert, W., Horvitz, H.R., Gunnarsson, L.-G., Brown, R.H., 2006. A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. Neurology 66, 839–844.

Morris, M.J., Wingate, K.L., Silwal, J., Leeper, T.C., Basu, S., 2012. The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. Nucleic Acids Res. 40, 4137–45. doi:10.1093/nar/gkr1308

Mulders, S.A.M., van den Broek, W.J.A.A., Wheeler, T.M., Croes, H.J.E., van Kuik-Romeijn, P., de Kimpe, S.J., Furling, D., Platenburg, G.J., Gourdon, G., Thornton, C.A., Wieringa, B., Wansink, D.G., 2009. Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy. Proc. Natl. Acad. Sci. U. S. A. 106, 13915–20. doi:10.1073/pnas.0905780106

Neary, D., Snowden, J., Mann, D., 2005. Frontotemporal dementia. Lancet Neurol. 4, 771–780. doi:10.1016/S1474-4422(05)70223-4

Neary, D., Snowden, J.S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P.H., Albert, M., Boone, K., Miller, B.L., Cummings, J., Benson, D.F., 1998. Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. Neurology 51, 1546–1554. doi:10.1212/WNL.51.6.1546

Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M., McCluskey, L.F., Miller, B.L., Masliah, E., Mackenzie, I.R., Feldman, H., Feiden, W., Kretzschmar, H.A., Trojanowski, J.Q., Lee, V.M.-Y., 2006. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science (80-.). 314, 130–133.

Nishimura, A.L., Mitne-Neto, M., Silva, H.C.A., Richieri-Costa, A., Middleton, S., Cascio, D., Kok, F., Oliveira, J.R.M., Gillingwater, T., Webb, J., Skehel, P., Zatz, M., 2004. A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. Am. J. Hum. Genet. 75, 822–31. doi:10.1086/425287

Nishimura, A.L., Zupunski, V., Troakes, C., Kathe, C., Fratta, P., Howell, M., Gallo, J.-M., Hortobágyi, T., Shaw, C.E., Rogelj, B., 2010. Nuclear import impairment causes cytoplasmic trans-activation response DNA-binding protein accumulation and is associated with frontotemporal lobar degeneration. Brain 133, 1763–71. doi:10.1093/brain/awq111

O'Rourke, J.G., Bogdanik, L., Muhammad, A.K.M.G., Gendron, T.F., Kim, K.J., Austin, A., Cady, J., Liu, E.Y., Zarrow, J., Grant, S., Ho, R., Bell, S., Carmona, S., Simpkinson, M., Lall, D., Wu, K., Daughrity, L., Dickson, D.W., Harms, M.B., Petrucelli, L., Lee, E.B., Lutz, C.M., Baloh, R.H., 2015. C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD. Neuron 88, 892–901. doi:10.1016/j.neuron.2015.10.027

Olson, M.O.J., 2011. The Nucleolus. Springer Science & Business Media.

Olson, M.O.J., 2004. Sensing cellular stress: another new function for the nucleolus? Sci. STKE 2004, pe10. doi:10.1126/stke.2242004pe10

Orengo, J.P., Chambon, P., Metzger, D., Mosier, D.R., Snipes, G.J., Cooper, T.A., 2008. Expanded CTG repeats within the DMPK 3' UTR causes severe skeletal muscle wasting in an inducible mouse model for myotonic dystrophy. Proc. Natl. Acad. Sci. U. S. A. 105, 2646–2651.

Parkinson, N., Ince, P.G., Smith, M.O., Highley, R., Skibinski, G., Andersen, P.M., Morrison, K.E., Pall, H.S., Hardiman, O., Collinge, J., Shaw, P.J., Fisher, E.M.C., 2006. ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B). Neurology 67, 1074–7. doi:10.1212/01.wnl.0000231510.89311.8b

Parlato, R., Kreiner, G., 2013. Nucleolar activity in neurodegenerative diseases: a missing piece of the puzzle? J. Mol. Med. (Berl). 91, 541–7. doi:10.1007/s00109-012-0981-1

Parlato, R., Kreiner, G., Erdmann, G., Rieker, C., Stotz, S., Savenkova, E., Berger, S., Grummt, I., Schütz, G., 2008. Activation of an endogenous suicide response after perturbation of rRNA synthesis leads to neurodegeneration in mice. J. Neurosci. 28, 12759–64. doi:10.1523/JNEUROSCI.2439-08.2008

Parlato, R., Liss, B., 2014. How Parkinson's disease meets nucleolar stress. Biochim. Biophys. Acta - Mol. Basis Dis. 1842, 791–797. doi:10.1016/j.bbadis.2013.12.014

Parniewski, P., Jaworski, A., Wells, R.D., Bowater, R.P., 2000. Length of CTG.CAG repeats determines the influence of mismatch repair on genetic instability. J. Mol. Biol. 299, 865–874. doi:10.1006/jmbi.2000.3796

Pasinelli, P., Brown, R.H., 2006. Molecular biology of amyotrophic lateral sclerosis: insights from genetics. Nat. Rev. Neurosci. 7, 710–23. doi:10.1038/nrn1971

Peabody, D.S., 1989. Translation initiation at non-AUG triplets in mammalian cells. J. Biol. Chem. 264, 5031–5035.

Pearson, J.P., Williams, N.M., Majounie, E., Waite, A., Stott, J., Newsway, V., Murray, A., Hernandez, D., Guerreiro, R., Singleton, A.B., Neal, J., Morris, H.R., 2011. Familial frontotemporal dementia with amyotrophic lateral sclerosis and a shared haplotype on chromosome 9p. J. Neurol. 258, 647–55. doi:10.1007/s00415-010-5815-x

Penagarikano, O., Mulle, J.G., Warren, S.T., 2007. The pathophysiology of fragile x syndrome. Annu. Rev. Genomics Hum. Genet. 8, 109–29. doi:10.1146/annurev.genom.8.080706.092249

Peters, O.M., Cabrera, G.T., Tran, H., Gendron, T.F., McKeon, J.E., Metterville, J., Weiss, A., Wightman, N., Salameh, J., Kim, J., Sun, H., Boylan, K.B., Dickson, D., Kennedy, Z., Lin, Z., Zhang, Y.-J., Daughrity, L., Jung, C., Gao, F.-B., Sapp, P.C., Horvitz, H.R., Bosco, D.A., Brown, S.P., de Jong, P., Petrucelli, L., Mueller, C., Brown, R.H., 2015. Human C9ORF72 Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but

Not Neurodegeneration in BAC Transgenic Mice. Neuron 88, 902–9. doi:10.1016/j.neuron.2015.11.018

Pieretti, M., Zhang, F., Fu, Y.-H., Warren, S.T., Oostra, B.A., Caskey, C.T., Nelson, D.L., 1991. Absence of expression of the FMR-1 gene in fragile X syndrome. Cell 66, 817–822. doi:10.1016/0092-8674(91)90125-I

Polymenidou, M., Lagier-Tourenne, C., Hutt, K.R., Huelga, S.C., Moran, J., Liang, T.Y., Ling, S.-C., Sun, E., Wancewicz, E., Mazur, C., Kordasiewicz, H., Sedaghat, Y., Donohue, J.P., Shiue, L., Bennett, C.F., Yeo, G.W., Cleveland, D.W., 2011. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat. Neurosci. 14, 459–68. doi:10.1038/nn.2779

Proudfoot, M., Gutowski, N.J., Edbauer, D., Hilton, D.A., Stephens, M., Rankin, J., Mackenzie, I.R.A., 2014. Early dipeptide repeat pathology in a frontotemporal dementia kindred with C9ORF72 mutation and intellectual disability. Acta Neuropathol. 127, 451–8. doi:10.1007/s00401-014-1245-7

Prudencio, M., Belzil, V. V, Batra, R., Ross, C.A., Gendron, T.F., Pregent, L.J., Murray, M.E., Overstreet, K.K., Piazza-Johnston, A.E., Desaro, P., Bieniek, K.F., DeTure, M., Lee, W.C., Biendarra, S.M., Davis, M.D., Baker, M.C., Perkerson, R.B., van Blitterswijk, M., Stetler, C.T., Rademakers, R., Link, C.D., Dickson, D.W., Boylan, K.B., Li, H., Petrucelli, L., 2015. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. Nat. Neurosci. 18, 1175–1182. doi:10.1038/nn.4065

Rademakers, R., Cruts, M., van Broeckhoven, C., 2004. The role of tau (MAPT) in frontotemporal dementia and related tauopathies. Hum. Mutat. 24, 277–295. doi:10.1002/humu.20086

Rademakers, R., Neumann, M., Mackenzie, I.R., 2012. Advances in understanding the molecular basis of frontotemporal dementia. Nat. Rev. Neurol. 8, 1–12. doi:10.1038/nrneurol.2012.117

Ratnavalli, E., Brayne, C., Dawson, K., Hodges, J.R., 2002. The prevalence of frontotemporal dementia. Neurology 58, 1615–21.

Ratti, A., Corrado, L., Castellotti, B., Del Bo, R., Fogh, I., Cereda, C., Tiloca, C., D'Ascenzo, C., Bagarotti, A., Pensato, V., Ranieri, M., Gagliardi, S., Calini, D., Mazzini, L., Taroni, F., Corti, S., Ceroni, M., Oggioni, G.D., Lin, K., Powell, J.F., Sorarù, G., Ticozzi, N., Comi, G.P., D'Alfonso, S., Gellera, C., Silani, V., 2012. C9ORF72 repeat expansion in a large Italian ALS cohort: evidence of a founder effect. Neurobiol. Aging 33, 2528.e7–14. doi:10.1016/j.neurobiolaging.2012.06.008

Reddy, K., Zamiri, B., Stanley, S.Y.R., Macgregor, R.B., Pearson, C.E., 2013. The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. J. Biol. Chem. 288, 9860–6. doi:10.1074/jbc.C113.452532

Renton, A., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, Schymick, J.,

Laaksovirta, H., Van Swieten, J., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A., Kaganovich, A., Scholz, S., Duckworth, J., Ding, J., Harmer, D., Hernandez, D., Johnson, J., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R., Orrell, R., Neal, J., Murray, A., Pearson, J., Jansen, I., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M., Peuralinna, T., Jansson, L., Isoviita, V., Kaivorinne, A., Holtta-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chio, A., Restagno, G., Borghero, G., Sabatelli, M., Heckerman, D., Rogaeva, E., Zinman, L., Rothstein, J., Sendtner, M., Drepper, C., Eichler, E., Alkan, C., Abdullaev, Z., Pack, S., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N., Heutink, P., Pickering-Brown, S., Morris, H., Tienari, P., Traynor, B., Consortium, I., 2011a. A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. Neuron 72, 1–12.

Renton, A.E., Chiò, A., Traynor, B.J., 2014. State of play in amyotrophic lateral sclerosis genetics. Nat. Neurosci. 17, 17–23. doi:10.1038/nn.3584

Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A.M., Kaganovich, A., Scholz, S.W., Duckworth, J., Ding, J., Harmer, D.W., Hernandez, D.G., Johnson, J.O., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R.J., Orrell, R.W., Neal, J., Murray, A., Pearson, J., Jansen, I.E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J.B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M. a, Peuralinna, T., Jansson, L., Isoviita, V.-M., Kaivorinne, A.-L., Hölttä-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chiò, A., Restagno, G., Borghero, G., Sabatelli, M., Heckerman, D., Rogaeva, E., Zinman, L., Rothstein, J.D., Sendtner, M., Drepper, C., Eichler, E.E., Alkan, C., Abdullaev, Z., Pack, S.D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N.M., Heutink, P., Pickering-Brown, S., Morris, H.R., Tienari, P.J., Traynor, B.J., 2011b. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72, 257–68. doi:10.1016/j.neuron.2011.09.010

Ringholz, G.M., Appel, S.H., Bradshaw, M., Cooke, N.A., Mosnik, D.M., Schulz, P.E., 2005. Prevalence and patterns of cognitive impairment in sporadic ALS. Neurology 65, 586–90. doi:10.1212/01.wnl.0000172911.39167.b6

Rohrer, J.D., Guerreiro, R., Vandrovcova, J., Uphill, J., Reiman, D., Beck, J., Isaacs, A.M., Authier, A., Ferrari, R., Fox, N.C., Mackenzie, I.R.A., Warren, J.D., de Silva, R., Holton, J., Revesz, T., Hardy, J., Mead, S., Rossor, M.N., 2009. The heritability and genetics of frontotemporal lobar degeneration. Neurology 73, 1451–6. doi:10.1212/WNL.0b013e3181bf997a

Rosen, D.R., Siddique, T., Patterson, D., Figlewicz, D.A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J.P., Deng, H.X., 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature 362, 59–62. doi:10.1038/362059a0

Rossi, S., Serrano, a., Gerbino, V., Giorgi, a., Di Francesco, L., Nencini, M., Bozzo, F., Schinina, M.E., Bagni, C., Cestra, G., Carri, M.T., Achsel, T., Cozzolino, M., 2015. Nuclear accumulation of mRNAs underlies G4C2-repeat-induced translational repression in a cellular model of C9orf72 ALS. J. Cell Sci. 128, 1787–1799. doi:10.1242/jcs.165332

Russ, J., Liu, E.Y., Wu, K., Neal, D., Suh, E., Irwin, D.J., McMillan, C.T., Harms, M.B., Cairns, N.J., Wood, E.M., Xie, S.X., Elman, L., McCluskey, L., Grossman, M., Van Deerlin, V.M., Lee, E.B., 2015. Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. Acta Neuropathol. 129, 39–52. doi:10.1007/s00401-014-1365-0

Sakamoto, N., Larson, J.E., Iyer, R.R., Montermini, L., Pandolfo, M., Wells, R.D., 2001. GGA*TCC-interrupted triplets in long GAA*TTC repeats inhibit the formation of triplex and sticky DNA structures, alleviate transcription inhibition, and reduce genetic instabilities. J. Biol. Chem. 276, 27178–27187. doi:10.1074/jbc.M101852200

Sareen, D., O'Rourke, J.G., Meera, P., Muhammad, A.K.M.G., Grant, S., Simpkinson, M., Bell, S., Carmona, S., Ornelas, L., Sahabian, A., Gendron, T., Petrucelli, L., Baughn, M., Ravits, J., Harms, M.B., Rigo, F., Bennett, C.F., Otis, T.S., Svendsen, C.N., Baloh, R.H., 2013. Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. Sci. Transl. Med. 5, 208ra149. doi:10.1126/scitranslmed.3007529

Saxena, S., Caroni, P., 2011. Selective neuronal vulnerability in neurodegenerative diseases: from stressor thresholds to degeneration. Neuron 71, 35–48. doi:10.1016/j.neuron.2011.06.031

Schludi, M.H., May, S., Grässer, F.A., Rentzsch, K., Kremmer, E., Küpper, C., Klopstock, T., Arzberger, T., Edbauer, D., 2015. Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. Acta Neuropathol. doi:10.1007/s00401-015-1450-z

Seelaar, H., Kamphorst, W., Rosso, S.M., Azmani, A., Masdjedi, R., de Koning, I., Maat-Kievit, J.A., Anar, B., Donker Kaat, L., Breedveld, G.J., Dooijes, D., Rozemuller, J.M., Bronner, I.F., Rizzu, P., van Swieten, J.C., 2008. Distinct genetic forms of frontotemporal dementia. Neurology 71, 1220–6. doi:10.1212/01.wnl.0000319702.37497.72

Sha, S.J., Takada, L.T., Rankin, K.P., Yokoyama, J.S., Rutherford, N.J., Fong, J.C., Khan, B., Karydas, A., Baker, M.C., DeJesus-Hernandez, M., Pribadi, M., Coppola, G., Geschwind, D.H., Rademakers, R., Lee, S.E., Seeley, W., Miller, B.L., Boxer, A.L., 2012. Frontotemporal dementia due to C9ORF72 mutations: clinical and imaging features. Neurology 79, 1002–11. doi:10.1212/WNL.0b013e318268452e

Shatunov, A., Mok, K., Newhouse, S., Weale, M.E., Smith, B., Vance, C., Johnson, L., Veldink, J.H., van Es, M.A., van den Berg, L.H., Robberecht, W., Van Damme, P., Hardiman, O., Farmer, A.E., Lewis, C.M., Butler, A.W., Abel, O., Andersen, P.M., Fogh, I., Silani, V., Chiò, A., Traynor, B.J., Melki, J., Meininger, V., Landers, J.E., McGuffin, P., Glass, J.D.,

Pall, H., Leigh, P.N., Hardy, J., Brown, R.H., Powell, J.F., Orrell, R.W., Morrison, K.E., Shaw, P.J., Shaw, C.E., Al-Chalabi, A., 2010. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. Lancet. Neurol. 9, 986–94. doi:10.1016/S1474-4422(10)70197-6

Shaw, G., Morse, S., Ararat, M., Graham, F.L., 2002. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. FASEB J. 16, 869–71. doi:10.1096/fj.01-0995fje

Sieben, A., Van Langenhove, T., Engelborghs, S., Martin, J.-J., Boon, P., Cras, P., De Deyn, P.-P., Santens, P., Van Broeckhoven, C., Cruts, M., 2012. The genetics and neuropathology of frontotemporal lobar degeneration. Acta Neuropathol. 124, 353–72. doi:10.1007/s00401-012-1029-x

Simón-Sánchez, J., Dopper, E.G.P., Cohn-Hokke, P.E., Hukema, R.K., Nicolaou, N., Seelaar, H., de Graaf, J.R.A., de Koning, I., van Schoor, N.M., Deeg, D.J.H., Smits, M., Raaphorst, J., van den Berg, L.H., Schelhaas, H.J., De Die-Smulders, C.E.M., Majoor-Krakauer, D., Rozemuller, A.J.M., Willemsen, R., Pijnenburg, Y.A.L., Heutink, P., van Swieten, J.C., 2012. The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions. Brain 135, 723–35. doi:10.1093/brain/awr353

Skibinski, G., Parkinson, N.J., Brown, J.M., Chakrabarti, L., Lloyd, S.L., Hummerich, H., Nielsen, J.E., Hodges, J.R., Spillantini, M.G., Thusgaard, T., Brandner, S., Brun, A., Rossor, M.N., Gade, A., Johannsen, P., Sørensen, S.A., Gydesen, S., Fisher, E.M.C., Collinge, J., 2005. Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia. Nat. Genet. 37, 806–8. doi:10.1038/ng1609

Smith, B.N., Newhouse, S., Shatunov, A., Vance, C., Topp, S., Johnson, L., Miller, J., Lee, Y., Troakes, C., Scott, K.M., Jones, A., Gray, I., Wright, J., Hortobágyi, T., Al-Sarraj, S., Rogelj, B., Powell, J., Lupton, M., Lovestone, S., Sapp, P.C., Weber, M., Nestor, P.J., Schelhaas, H.J., Asbroek, A.A. Ten, Silani, V., Gellera, C., Taroni, F., Ticozzi, N., Van den Berg, L., Veldink, J., Van Damme, P., Robberecht, W., Shaw, P.J., Kirby, J., Pall, H., Morrison, K.E., Morris, A., de Belleroche, J., Vianney de Jong, J.M.B., Baas, F., Andersen, P.M., Landers, J., Brown, R.H., Weale, M.E., Al-Chalabi, A., Shaw, C.E., 2013. The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. Eur. J. Hum. Genet. 21, 102–8. doi:10.1038/ejhg.2012.98

Snowden, J.S., Rollinson, S., Thompson, J.C., Harris, J.M., Stopford, C.L., Richardson, A.M.T., Jones, M., Gerhard, A., Davidson, Y.S., Robinson, A., Gibbons, L., Hu, Q., DuPlessis, D., Neary, D., Mann, D.M.A., Pickering-Brown, S.M., 2012. Distinct clinical and pathological characteristics of frontotemporal dementia associated with C9ORF72 mutations. Brain 135, 693–708. doi:10.1093/brain/awr355

Spillantini, M.G., Goedert, M., 2013. Tau pathology and neurodegeneration. Lancet Neurol. 12, 609–22. doi:10.1016/S1474-4422(13)70090-5

Stewart, H., Rutherford, N.J., Briemberg, H., Krieger, C., Cashman, N., Fabros, M., Baker, M., Fok, A., DeJesus-Hernandez, M., Eisen, A., Rademakers, R., Mackenzie, I.R.A., 2012.

Clinical and pathological features of amyotrophic lateral sclerosis caused by mutation in the C9ORF72 gene on chromosome 9p. Acta Neuropathol. 123, 409–417. doi:10.1007/s00401-011-0937-5

Stutz, F., Bachi, A., Doerks, T., Braun, I.C., Seraphin, B., Wilm, M., Bork, P., Izaurralde, E., 2000. REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. RNA 6, 638–650.

Su, Z., Zhang, Y., Gendron, T.F., Bauer, P.O., Chew, J., Yang, W.-Y., Fostvedt, E., Jansen-West, K., Belzil, V.V., Desaro, P., Johnston, A., Overstreet, K., Boeve, B.F., Dickson, D., Floeter, M.K., Traynor, B.J., Morelli, C., Ratti, A., Silani, V., Rademakers, R., Brown, R.H., Rothstein, J.D., Boylan, K.B., Petrucelli, L., Disney, M.D., 2014. Discovery of a Biomarker and Lead Small Molecules to Target r(GGGGCC)-Associated Defects in c9FTD/ALS. Neuron. doi:10.1016/j.neuron.2014.07.041

Suh, E., Lee, E.B., Neal, D., Wood, E.M., Toledo, J.B., Rennert, L., Irwin, D.J., McMillan, C.T., Krock, B., Elman, L.B., McCluskey, L.F., Grossman, M., Xie, S.X., Trojanowski, J.Q., Van Deerlin, V.M., 2015. Semi-automated quantification of C9orf72 expansion size reveals inverse correlation between hexanucleotide repeat number and disease duration in frontotemporal degeneration. Acta Neuropathol. doi:10.1007/s00401-015-1445-9

Sutcliffe, J.S., Nelson, D.L., Zhang, F., Pieretti, M., Caskey, C.T., Saxe, D., Warren, S.T., 1992. DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum. Mol. Genet. 1, 397–400. doi:10.1093/hmg/1.6.397

Sutedja, N.A., Fischer, K., Veldink, J.H., Heijden, G.J.M.G. van der, Kromhout, H., Heederik, D., Huisman, M.H.B., Wokke, J.J.H., Berg, L.H. van den, 2009a. What we truly know about occupation as a risk factor for ALS: A critical and systematic review. Amyotroph. Lateral Scler.

Sutedja, N.A., Veldink, J.H., Fischer, K., Kromhout, H., Heederik, D., Huisman, M.H.B., Wokke, J.H.J., van den Berg, L.H., 2009b. Exposure to chemicals and metals and risk of amyotrophic lateral sclerosis: a systematic review. Amyotroph. Lateral Scler. 10, 302–9. doi:10.3109/17482960802455416

Taneja, K.L., 1995. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. J. Cell Biol. 128, 995–1002. doi:10.1083/jcb.128.6.995

Tao, Z., Wang, H., Xia, Q., Li, K., Jiang, X., Xu, G., Wang, G., Ying, Z., 2015. Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. Hum. Mol. Genet. ddv005–. doi:10.1093/hmg/ddv005

Tatom, J.B., Wang, D.B., Dayton, R.D., Skalli, O., Hutton, M.L., Dickson, D.W., Klein, R.L., 2009. Mimicking aspects of frontotemporal lobar degeneration and Lou Gehrig's disease in rats via TDP-43 overexpression. Mol. Ther. 17, 607–13. doi:10.1038/mt.2009.3

Therrien, M., Rouleau, G.A., Dion, P.A., Parker, J.A., 2013. Deletion of C9ORF72 results in motor neuron degeneration and stress sensitivity in C. elegans. PLoS One 8, e83450.

doi:10.1371/journal.pone.0083450

Theuns, J., Verstraeten, A., Sleegers, K., Wauters, E., Gijselinck, I., Smolders, S., Crosiers, D., Corsmit, E., Elinck, E., Sharma, M., Krüger, R., Lesage, S., Brice, A., Chung, S.J., Kim, M.-J., Kim, Y.J., Ross, O.A., Wszolek, Z.K., Rogaeva, E., Xi, Z., Lang, A.E., Klein, C., Weissbach, A., Mellick, G.D., Silburn, P.A., Hadjigeorgiou, G.M., Dardiotis, E., Hattori, N., Ogaki, K., Tan, E.-K., Zhao, Y., Aasly, J., Valente, E.M., Petrucci, S., Annesi, G., Quattrone, A., Ferrarese, C., Brighina, L., Deutschländer, A., Puschmann, A., Nilsson, C., Garraux, G., LeDoux, M.S., Pfeiffer, R.F., Boczarska-Jedynak, M., Opala, G., Maraganore, D.M., Engelborghs, S., De Deyn, P.P., Cras, P., Cruts, M., Van Broeckhoven, C., 2014. Global investigation and meta-analysis of the C9orf72 (G4C2)n repeat in Parkinson disease. Neurology 83, 1906–13. doi:10.1212/WNL.0000000000001012

Thornton, C.A., Wymer, J.P., Simmons, Z., McClain, C., Moxley, R.T., 1997. Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. Nat. Genet. 16, 407–9. doi:10.1038/ng0897-407

Todd, P.K., Oh, S.Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A.J., Chen, K., Scaglione, K.M., Basrur, V., Elenitoba-Johnson, K., Vonsattel, J.P., Louis, E.D., Sutton, M.A., Taylor, J.P., Mills, R.E., Charlet-Berguerand, N., Paulson, H.L., 2013. CGG Repeat-Associated Translation Mediates Neurodegeneration in Fragile X Tremor Ataxia Syndrome. Neuron 78, 440–455. doi:10.1016/j.neuron.2013.03.026

Tran, H., Almeida, S., Moore, J., Gendron, T.F., Chalasani, U., Lu, Y., Du, X., Nickerson, J.A., Petrucelli, L., Weng, Z., Gao, F.-B., 2015. Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. Neuron 87, 1207–14. doi:10.1016/j.neuron.2015.09.015

Tsoi, H., Chan, H.Y.E., 2014. Roles of the nucleolus in the CAG RNA-mediated toxicity. Biochim. Biophys. Acta 1842, 779–84. doi:10.1016/j.bbadis.2013.11.015

Valdmanis, P.N., Dupre, N., Bouchard, J.-P., Camu, W., Salachas, F., Meininger, V., Strong, M., Rouleau, G.A., 2007. Three families with amyotrophic lateral sclerosis and frontotemporal dementia with evidence of linkage to chromosome 9p. Arch. Neurol. 64, 240–5. doi:10.1001/archneur.64.2.240

van Blitterswijk, M., DeJesus-Hernandez, M., Niemantsverdriet, E., Murray, M.E., Heckman, M.G., Diehl, N.N., Brown, P.H., Baker, M.C., Finch, N.A., Bauer, P.O., Serrano, G., Beach, T.G., Josephs, K.A., Knopman, D.S., Petersen, R.C., Boeve, B.F., Graff-Radford, N.R., Boylan, K.B., Petrucelli, L., Dickson, D.W., Rademakers, R., 2013. Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. Lancet Neurol. 12, 978–88. doi:10.1016/S1474-4422(13)70210-2

van Blitterswijk, M., Gendron, T.F., Baker, M.C., DeJesus-Hernandez, M., Finch, N.A., Brown, P.H., Daughrity, L.M., Murray, M.E., Heckman, M.G., Jiang, J., Lagier-Tourenne, C., Edbauer, D., Cleveland, D.W., Josephs, K.A., Parisi, J.E., Knopman, D.S., Petersen, R.C.,

Petrucelli, L., Boeve, B.F., Graff-Radford, N.R., Boylan, K.B., Dickson, D.W., Rademakers, R., 2015. Novel clinical associations with specific C9ORF72 transcripts in patients with repeat expansions in C9ORF72. Acta Neuropathol. doi:10.1007/s00401-015-1480-6

Van Deerlin, V.M., Sleiman, P.M.A., Martinez-Lage, M., Chen-Plotkin, A., Wang, L.-S., Graff-Radford, N.R., Dickson, D.W., Rademakers, R., Boeve, B.F., Grossman, M., Arnold, S.E., Mann, D.M.A., Pickering-Brown, S.M., Seelaar, H., Heutink, P., van Swieten, J.C., Murrell, J.R., Ghetti, B., Spina, S., Grafman, J., Hodges, J., Spillantini, M.G., Gilman, S., Lieberman, A.P., Kaye, J.A., Woltjer, R.L., Bigio, E.H., Mesulam, M., Al-Sarraj, S., Troakes, C., Rosenberg, R.N., White, C.L., Ferrer, I., Lladó, A., Neumann, M., Kretzschmar, H.A., Hulette, C.M., Welsh-Bohmer, K.A., Miller, B.L., Alzualde, A., Lopez de Munain, A., McKee, A.C., Gearing, M., Levey, A.I., Lah, J.J., Hardy, J., Rohrer, J.D., Lashley, T., Mackenzie, I.R.A., Feldman, H.H., Hamilton, R.L., Dekosky, S.T., van der Zee, J., Kumar-Singh, S., Van Broeckhoven, C., Mayeux, R., Vonsattel, J.P.G., Troncoso, J.C., Kril, J.J., Kwok, J.B.J., Halliday, G.M., Bird, T.D., Ince, P.G., Shaw, P.J., Cairns, N.J., Morris, J.C., McLean, C.A., DeCarli, C., Ellis, W.G., Freeman, S.H., Frosch, M.P., Growdon, J.H., Perl, D.P., Sano, M., Bennett, D.A., Schneider, J.A., Beach, T.G., Reiman, E.M., Woodruff, B.K., Cummings, J., Vinters, H. V, Miller, C.A., Chui, H.C., Alafuzoff, I., Hartikainen, P., Seilhean, D., Galasko, D., Masliah, E., Cotman, C.W., Tuñón, M.T., Martínez, M.C.C., Munoz, D.G., Carroll, S.L., Marson, D., Riederer, P.F., Bogdanovic, N., Schellenberg, G.D., Hakonarson, H., Trojanowski, J.Q., Lee, V.M.-Y., 2010. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nat. Genet. 42, 234–9. doi:10.1038/ng.536

van Der Laak, J.A., Pahlplatz, M.M., Hanselaar, A.G., de Wilde, P.C., 2000. Hue-saturation-density (HSD) model for stain recognition in digital images from transmitted light microscopy. Cytometry 39, 275–84.

van der Zee, J., Gijselinck, I., Dillen, L., Van Langenhove, T., Theuns, J., Engelborghs, S., Philtjens, S., Vandenbulcke, M., Sleegers, K., Sieben, A., Bäumer, V., Maes, G., Corsmit, E., Borroni, B., Padovani, A., Archetti, S., Perneczky, R., Diehl-Schmid, J., de Mendonça, A., Miltenberger-Miltenyi, G., Pereira, S., Pimentel, J., Nacmias, B., Bagnoli, S., Sorbi, S., Graff, C., Chiang, H.-H., Westerlund, M., Sanchez-Valle, R., Llado, A., Gelpi, E., Santana, I., Almeida, M.R., Santiago, B., Frisoni, G., Zanetti, O., Bonvicini, C., Synofzik, M., Maetzler, W., Vom Hagen, J.M., Schöls, L., Heneka, M.T., Jessen, F., Matej, R., Parobkova, E., Kovacs, G.G., Ströbel, T., Sarafov, S., Tournev, I., Jordanova, A., Danek, A., Arzberger, T., Fabrizi, G.M., Testi, S., Salmon, E., Santens, P., Martin, J.-J., Cras, P., Vandenberghe, R., De Deyn, P.P., Cruts, M., Van Broeckhoven, C., De Deyn, P.P., Müller Vom Hagen, J., Ramirez, A., Kurzwelly, D., Sachtleben, C., Mairer, W., Firmo, C., Antonell, A., Molinuevo, J., Kinhult Ståhlbom, A., Thonberg, H., Nennesmo, I., Börjesson-Hanson, A., Bessi, V., Piaceri, I., Helena Ribeiro, M., Rosário Almeida, M., Oliveira, C., Massano, J., Garret, C., Pires, P., Danel, A., Maria Fabrizi, G., Ferrari, S., Cavallaro, T., 2013. A pan-European study of the C9orf72 repeat associated with FTLD: geographic prevalence, genomic instability, and intermediate repeats. Hum. Mutat.

34, 363–73. doi:10.1002/humu.22244

van Es, M.A., Veldink, J.H., Saris, C.G.J., Blauw, H.M., van Vught, P.W.J., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J.N., Huisman, M.H.B., van der Kooi, A.J., de Visser, M., Dahlberg, C., Estrada, K., Rivadeneira, F., Hofman, A., Zwarts, M.J., van Doormaal, P.T.C., Rujescu, D., Strengman, E., Giegling, I., Muglia, P., Tomik, B., Slowik, A., Uitterlinden, A.G., Hendrich, C., Waibel, S., Meyer, T., Ludolph, A.C., Glass, J.D., Purcell, S., Cichon, S., Nöthen, M.M., Wichmann, H.-E., Schreiber, S., Vermeulen, S.H.H.M., Kiemeney, L.A., Wokke, J.H.J., Cronin, S., McLaughlin, R.L., Hardiman, O., Fumoto, K., Pasterkamp, R.J., Meininger, V., Melki, J., Leigh, P.N., Shaw, C.E., Landers, J.E., Al-Chalabi, A., Brown, R.H., Robberecht, W., Andersen, P.M., Ophoff, R.A., van den Berg, L.H., 2009. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat. Genet. 41, 1083–7. doi:10.1038/ng.442

Van Langenhove, T., van der Zee, J., Gijselinck, I., Engelborghs, S., Vandenberghe, R., Vandenbulcke, M., De Bleecker, J., Sieben, A., Versijpt, J., Ivanoiu, A., Deryck, O., Willems, C., Dillen, L., Philtjens, S., Maes, G., Bäumer, V., Van Den Broeck, M., Mattheijssens, M., Peeters, K., Martin, J.-J., Michotte, A., Santens, P., De Jonghe, P., Cras, P., De Deyn, P.P., Cruts, M., Van Broeckhoven, C., 2013. Distinct Clinical Characteristics of C9orf72 Expansion Carriers Compared With GRN, MAPT, and Nonmutation Carriers in a Flanders-Belgian FTLD Cohort. JAMA Neurol. 70, 365. doi:10.1001/2013.jamaneurol.181

Van Langenhove, T., van der Zee, J., Sleegers, K., Engelborghs, S., Vandenberghe, R., Gijselinck, I., Van den Broeck, M., Mattheijssens, M., Peeters, K., De Deyn, P.P., Cruts, M., Van Broeckhoven, C., 2010. Genetic contribution of FUS to frontotemporal lobar degeneration. Neurology 74, 366–71. doi:10.1212/WNL.0b013e3181ccc732

Van Langenhove, T., van der Zee, J., Van Broeckhoven, C., 2012. The molecular basis of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum. Ann. Med. 44, 817–28. doi:10.3109/07853890.2012.665471

Vance, C., Al-Chalabi, A., Ruddy, D., Smith, B.N., Hu, X., Sreedharan, J., Siddique, T., Schelhaas, H.J., Kusters, B., Troost, D., Baas, F., de Jong, V., Shaw, C.E., 2006. Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. Brain 129, 868–76. doi:10.1093/brain/awl030

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., Ganesalingam, J., Williams, K.L., Tripathi, V., Al-Saraj, S., Al-Chalabi, A., Leigh, P.N., Blair, I.P., Nicholson, G., de Belleroche, J., Gallo, J.-M., Miller, C.C., Shaw, C.E., 2009. Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. Science (80-.). 323, 1208–1211. doi:10.1126/science.1165942

Vossel, K.A., Miller, B.L., 2008. New approaches to the treatment of frontotemporal lobar degeneration. Curr. Opin. Neurol. 21, 708–16. doi:10.1097/WCO.0b013e328318444d

Waite, A.J., Bäumer, D., East, S., Neal, J., Morris, H.R., Ansorge, O., Blake, D.J., 2014. Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. Neurobiol. Aging 35, 1779.e5–1779.e13. doi:10.1016/j.neurobiolaging.2014.01.016

Ward, M.E., Taubes, A., Chen, R., Miller, B.L., Sephton, C.F., Gelfand, J.M., Minami, S., Boscardin, J., Martens, L.H., Seeley, W.W., Yu, G., Herz, J., Filiano, A.J., Arrant, A.E., Roberson, E.D., Kraft, T.W., Farese, R. V, Green, A., Gan, L., 2014. Early retinal neurodegeneration and impaired Ran-mediated nuclear import of TDP-43 in progranulin-deficient FTLD. J. Exp. Med. 211, 1937–45. doi:10.1084/jem.20140214

Watts, G.D.J., Wymer, J., Kovach, M.J., Mehta, S.G., Mumm, S., Darvish, D., Pestronk, A., Whyte, M.P., Kimonis, V.E., 2004. Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. Nat. Genet. 36, 377–81. doi:10.1038/ng1332

Weill, L., Belloc, E., Bava, F.-A., Méndez, R., 2012. Translational control by changes in poly(A) tail length: recycling mRNAs. Nat. Struct. Mol. Biol. 19, 577–85. doi:10.1038/nsmb.2311

Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S.S., Shi, Y., Lin, S., Shneider, N.A., Monaghan, J., Pandey, U.B., Pasinelli, P., Ichida, J.K., Trotti, D., 2014. Antisense Proline-Arginine RAN Dipeptides Linked to C9ORF72-ALS/FTD Form Toxic Nuclear Aggregates that Initiate In Vitro and In Vivo Neuronal Death. Neuron 84, 1213–1225. doi:10.1016/j.neuron.2014.12.010

Whitwell, J.L., Weigand, S.D., Boeve, B.F., Senjem, M.L., Gunter, J.L., DeJesus-Hernandez, M., Rutherford, N.J., Baker, M., Knopman, D.S., Wszolek, Z.K., Parisi, J.E., Dickson, D.W., Petersen, R.C., Rademakers, R., Jack, C.R., Josephs, K.A., 2012. Neuroimaging signatures of frontotemporal dementia genetics: C9ORF72, tau, progranulin and sporadics. Brain 135, 794–806. doi:10.1093/brain/aws001

Wils, H., Kleinberger, G., Janssens, J., Pereson, S., Joris, G., Cuijt, I., Smits, V., Ceuterick-de Groote, C., Van Broeckhoven, C., Kumar-Singh, S., 2010. TDP-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS and frontotemporal lobar degeneration. Proc. Natl. Acad. Sci. U. S. A. 107, 3858–3863.

Woerner, A.C., Frottin, F., Hornburg, D., Feng, L.R., Meissner, F., Patra, M., Tatzelt, J., Mann, M., Winklhofer, K.F., Hartl, F.U., Hipp, M.S., 2015. Cytoplasmic protein aggregates interfere with nucleo-cytoplasmic transport of protein and RNA. Science science.aad2033–. doi:10.1126/science.aad2033

Woollacott, I.O.C., Mead, S., 2014. The C9ORF72 expansion mutation: gene structure, phenotypic and diagnostic issues. Acta Neuropathol. 127, 319–32. doi:10.1007/s00401-014-1253-7

Wu, C.-H., Fallini, C., Ticozzi, N., Keagle, P.J., Sapp, P.C., Piotrowska, K., Lowe, P., Koppers,

M., McKenna-Yasek, D., Baron, D.M., Kost, J.E., Gonzalez-Perez, P., Fox, A.D., Adams, J., Taroni, F., Tiloca, C., Leclerc, A.L., Chafe, S.C., Mangroo, D., Moore, M.J., Zitzewitz, J.A., Xu, Z.-S., van den Berg, L.H., Glass, J.D., Siciliano, G., Cirulli, E.T., Goldstein, D.B., Salachas, F., Meininger, V., Rossoll, W., Ratti, A., Gellera, C., Bosco, D.A., Bassell, G.J., Silani, V., Drory, V.E., Brown, R.H., Landers, J.E., 2012. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. Nature 488, 499–503. doi:10.1038/nature11280

Xi, Z., Rainero, I., Rubino, E., Pinessi, L., Bruni, A.C., Maletta, R.G., Nacmias, B., Sorbi, S., Galimberti, D., Surace, E.I., Zheng, Y., Moreno, D., Sato, C., Liang, Y., Zhou, Y., Robertson, J., Zinman, L., Tartaglia, M.C., St George-Hyslop, P., Rogaeva, E., 2014. Hypermethylation of the CpG-island near the C9orf72 G_4C_2-repeat expansion in FTLD patients. Hum. Mol. Genet. 23, 5630–7. doi:10.1093/hmg/ddu279

Xi, Z., van Blitterswijk, M., Zhang, M., McGoldrick, P., McLean, J.R., Yunusova, Y., Knock, E., Moreno, D., Sato, C., McKeever, P.M., Schneider, R., Keith, J., Petrescu, N., Fraser, P., Tartaglia, M.C., Baker, M.C., Graff-Radford, N.R., Boylan, K.B., Dickson, D.W., Mackenzie, I.R., Rademakers, R., Robertson, J., Zinman, L., Rogaeva, E., 2015a. Jump from Pre-mutation to Pathologic Expansion in C9orf72. Am. J. Hum. Genet. 96, 962–970. doi:10.1016/j.ajhg.2015.04.016

Xi, Z., Zhang, M., Bruni, A.C., Maletta, R.G., Colao, R., Fratta, P., Polke, J.M., Sweeney, M.G., Mudanohwo, E., Nacmias, B., Sorbi, S., Tartaglia, M.C., Rainero, I., Rubino, E., Pinessi, L., Galimberti, D., Surace, E.I., McGoldrick, P., McKeever, P., Moreno, D., Sato, C., Liang, Y., Keith, J., Zinman, L., Robertson, J., Rogaeva, E., 2015b. The C9orf72 repeat expansion itself is methylated in ALS and FTLD patients. Acta Neuropathol. 129, 715–27. doi:10.1007/s00401-015-1401-8

Xi, Z., Zinman, L., Moreno, D., Schymick, J., Liang, Y., Sato, C., Zheng, Y., Ghani, M., Dib, S., Keith, J., Robertson, J., Rogaeva, E., 2013. Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. Am. J. Hum. Genet. 92, 981–9. doi:10.1016/j.ajhg.2013.04.017

Xiao, S., MacNair, L., McGoldrick, P., McKeever, P.M., McLean, J.R., Zhang, M., Keith, J., Zinman, L., Rogaeva, E., Robertson, J., 2015. Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis. Ann. Neurol. 78, 568–83. doi:10.1002/ana.24469

Xu, Z., Poidevin, M., Li, X., Li, Y., Shu, L., Nelson, D.L., Li, H., Hales, C.M., Gearing, M., Wingo, T.S., Jin, P., 2013. Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. Proc. Natl. Acad. Sci. 110, 7778–7783. doi:10.1073/pnas.1219643110

Yamakawa, M., Ito, D., Honda, T., Kubo, K.-I., Noda, M., Nakajima, K., Suzuki, N., 2014. Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. Hum. Mol. Genet. ddu576–. doi:10.1093/hmg/ddu576

Yang, D., Abdallah, A., Li, Z., Lu, Y., Almeida, S., Gao, F.-B., 2015. FTD/ALS-associated

poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. Acta Neuropathol. 130, 525–535. doi:10.1007/s00401-015-1448-6

Zamiri, B., Reddy, K., Macgregor, R.B., Pearson, C.E., 2014. TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. J. Biol. Chem. 289, 4653–9. doi:10.1074/jbc.C113.502336

Zhang, D., Iyer, L.M., He, F., Aravind, L., 2012. Discovery of Novel DENN Proteins: Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease. Front. Genet. 3, 283. doi:10.3389/fgene.2012.00283

Zhang, K., Donnelly, C.J., Haeusler, A.R., Grima, J.C., Machamer, J.B., Steinwald, P., Daley, E.L., Miller, S.J., Cunningham, K.M., Vidensky, S., Gupta, S., Thomas, M.A., Hong, I., Chiu, S.-L., Huganir, R.L., Ostrow, L.W., Matunis, M.J., Wang, J., Sattler, R., Lloyd, T.E., Rothstein, J.D., 2015. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. Nature. doi:10.1038/nature14973

Zhang, Y.-J., Jansen-West, K., Xu, Y.-F., Gendron, T.F., Bieniek, K.F., Lin, W.-L., Sasaguri, H., Caulfield, T., Hubbard, J., Daughrity, L., Chew, J., Belzil, V. V, Prudencio, M., Stankowski, J.N., Castanedes-Casey, M., Whitelaw, E., Ash, P.E.A., DeTure, M., Rademakers, R., Boylan, K.B., Dickson, D.W., Petrucelli, L., 2014. Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. Acta Neuropathol. doi:10.1007/s00401-014-1336-5

Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A.C., Nan, Z., Forster, C., Low, W.C., Schoser, B., Somia, N. V, Clark, H.B., Schmechel, S., Bitterman, P.B., Gourdon, G., Swanson, M.S., Moseley, M., Ranum, L.P.W., 2011. Non-ATG-initiated translation directed by microsatellite expansions. Proc. Natl. Acad. Sci. U. S. A. 108, 260–265. doi:10.1073/pnas.1013343108

Zu, T., Liu, Y., Bañez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T.M., Harms, M.B., Falchook, A.E., Subramony, S.H., Ostrow, L.W., Rothstein, J.D., Troncoso, J.C., Ranum, L.P.W., 2013. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc. Natl. Acad. Sci. U. S. A. 110, E4968–77. doi:10.1073/pnas.1315438110